

Exponential Strong Converse for Content Identification with Lossy Recovery

Lin Zhou, Vincent Y. F. Tan, and Mehul Motani

Department of Electrical and Computer Engineering,
National University of Singapore

Emails: lzhou@u.nus.edu, {vtan, motani}@nus.edu.sg

Abstract

In this paper, we revisit the high-dimensional content identification with lossy recovery problem (Tuncel and Gündüz, 2014). We first present a non-asymptotic converse bound. Invoking the non-asymptotic converse bound, we derive a lower bound on the exponent of the probability of correct decoding (the strong converse exponent) and show the lower bound is strictly positive if the rate-distortion tuple falls outside the rate-distortion region by Tuncel and Gündüz. Hence, we establish the exponential strong converse theorem for the content identification problem with lossy recovery. As corollaries of the exponential strong converse theorem, we derive an upper bound on the joint excess-distortion exponent for the problem. Our main results can be specialized to the biometrical identification problem (Willems, 2003) and the content identification problem (Tuncel, 2009) since these two problems are both special cases of the content identification problem with lossy recovery. We leverage the information spectrum method introduced by Oohama (2015, 2016). We adapt the strong converse techniques therein to be applicable to the problem at hand and we unify the analysis carefully to obtain the desired results.

I. INTRODUCTION

Have you ever wondered about the identity of a song after hearing only a short snippet? With limited information, it is sometimes difficult to identify the song or a distorted version of it, yet not impossible. In fact, there is an app called Shazam [1] that does precisely this. There are three distinct steps in the process of identifying the song, namely, the enrollment phase, the identification phase and the lossy recovery phase (see Figure 1). In the enrollment phase, the database of songs is sought; in the identification phase, we would like to infer certain details about the song; and finally, in the recovery phase, we hope to recover (at least) a lossy version of the song. An information-theoretic model was put forth by Tuncel and Gündüz [2] and they called this model the (high-dimensional) content identification problem with lossy recovery. This model is also applicable to other situations such as fingerprint identification [3] and video identification [4]. However, [2] only established a weak converse. In this paper, we revisit the content identification problem with lossy recovery and derive the exponential strong converse theorem.

A. Related Work

The most related works are [5] and [2]. In [5], Tuncel characterizes the achievable rate region of the content identification problem. In [2], Tuncel and Gündüz characterized the rate-distortion region for content identification problem with lossy recovery. Other (non-exhausting) works on the content identification problem are summarized as follows. Willems, Kalker, Goseling and Linnartz [6] initiated the study of the content identification problem by characterizing the capacity of a biometrical identification problem. Dasarathy and Draper [7] derived upper and lower bounds on the error exponent of the content identification system where they assume the DMC $P_{Y|X}$ is a noiseless channel. Recently, Merhav [8] refined the result in [7] by proposing a universal achievability scheme and showing that the scheme achieves the optimal exponent given by maximum likelihood decoding. Furthermore, Yachongka and Yagi [9] established the strong converse theorem for the biometrical identification problem. We remark that Yachongka and Yagi used Arimoto's strong converse technique [10] which is different from the information spectrum method adopted in this paper. The main result of [9] is recovered as a by-product of our main result. Other works on content identification include [11]–[17].

We also summarize the works by Oohama on using the information spectrum method to derive exponential strong converse theorems for several network information theory problems. In [18], [19] and [20] Oohama derived exponential strong converses for the one-helper (WAK) problem [21], [22], the asymmetric broadcast channel problem [23], and the Wyner-Ziv problem [24] respectively.

B. Main Contributions and Challenges

For the content identification problem with lossy recovery, we first present a non-asymptotic converse bound. Invoking the non-asymptotic converse bound, we establish an upper bound on the probability of correct decoding in both the content identification index and the feature vector. By correct decoding of the feature vector, we mean that the reproduced feature vector is within certain distortion level under a distortion measure. Further, we show that the probability of correct decoding

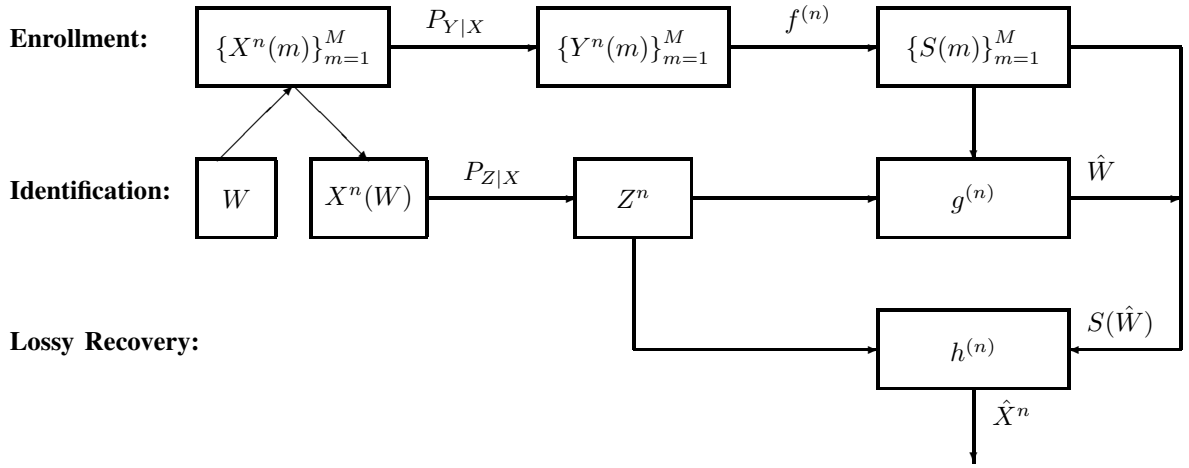


Fig. 1. Content identification with lossy recovery [2].

decays exponentially fast to zero if the rate-distortion tuple falls outside the rate-distortion region by Tuncel and Gündüz in [2]. Hence, we establish the exponential strong converse theorem for the content identification problem with lossy recovery. As a corollary, we derive an upper bound on the joint error and excess-distortion exponent. Our results can be specialized to the biometrical identification problem [6] and the content identification problem [5]. In particular, for the biometrical identification problem, we derive the moderate deviations constant and the second-order coding rate.

In the rest of this subsection, we discuss the main challenges in establishing the strong converse theorem for the content identification problem with lossy recovery. First, we need to identify the correct form of the auxiliary random variable. As can be seen in the proofs in Section IV, the auxiliary random variables we choose are different from those in the weak converse proof [2]. If we choose the auxiliary variable as in the weak converse proof, we cannot establish the exponential strong converse result.

Second, the content identification problem with lossy recovery involves three phases: the enrollment phase, the identification phase and the lossy recovery phase. It is challenging to unify the analyses in different stages since the same auxiliary random variable is shared in all phases. Hence, we adopt ideas from the strong converses in [19], [20] for the Wyner-Ziv problem and the degraded broadcast channel.

Third, in the identification phase, we need to use the whole random codebook and a noisy version of a certain feature vector to estimate the index of the feature vector (the identification index). This is very different from traditional channel coding and source coding problems. In source coding problems, we have only the codeword of a source sequence to decode the source while in channel coding problems, we have only the channel output for a particular message to decode. Hence, techniques like the image size characterization [25] and the perturbation approach [26] are probably insufficient to establish the strong converse for the current problem. As explained above, we adapt Oohama's strong converse techniques to deal with the challenges.

C. Organization of the Paper

The rest of the paper is organized as follows. In Section II, we set up the notation, formulate the content identification with lossy recovery problem and recapitulate the existing results concerning the rate-distortion region. In Section III, we first present a non-asymptotic upper bound on the probability of correct decoding and then claim the exponential strong converse by studying the properties of the bound. As a corollary, we derive an upper bound on the joint error and excess-distortion exponent. Our main results can be specialized to the content identification [5] and the biometrical identification [6] problems. Further, for the biometrical identification problem, we derive the moderate deviations constant and the second-order coding rate. The proof of our main result is presented in Section IV. Finally, we conclude the paper in Section V. For seamless presentation of results, the proofs of all supporting lemmas are deferred to the appendices.

II. PROBLEM FORMULATION AND EXISTING RESULTS

Notation

Random variables and their realizations are in capital (e.g., X) and lower case (e.g., x) respectively. All sets are denoted in calligraphic font (e.g., \mathcal{X}). We use \mathcal{X}^c to denote the complement of \mathcal{X} . Let $X^n := (X_1, \dots, X_n)$ be a random vector of length n . We use \mathbb{R}_+ to denote the set of positive real numbers. Given a number $a \in [0, 1]$, we use \bar{a} to denote $1 - a$. For quantities such as entropy and mutual information, we follow the notation in [25]. The set of all probability distribution on \mathcal{X} is denoted as $\mathcal{P}(\mathcal{X})$ and the set of all conditional probability distribution from \mathcal{X} to \mathcal{Y} is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X})$.

A. Problem Formulation

Let the random variables (X, Y, Z, \hat{X}) take values in finite alphabets \mathcal{X} , \mathcal{Y} , \mathcal{Z} and \mathcal{X} respectively. Let $\mathcal{M} := \{1, \dots, M\}$ and $\mathcal{L} := \{1, \dots, L\}$. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$ be the distortion measure and let the distortion between X^n and \hat{X}^n be defined as $d(X^n, \hat{X}^n) := \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i)$. Let the maximum distortion between $x \in \mathcal{X}$ and $\hat{x} \in \hat{\mathcal{X}}$ be d^+ , i.e., $d^+ := \max_{x, \hat{x}} d(x, \hat{x})$. Assume that each of the feature vectors $\{X^n(m)\}_{m \in \mathcal{M}}$ is generated i.i.d. according to P_X^n . The content identification problem with lossy recovery is divided into three phases: the enrollment phase, the identification phase and the lossy recovery phase. See Figure 1.

In the enrollment phase, for each $m \in \mathcal{M}$, the noisy version $Y^n(m)$ of each feature vector $X^n(m)$ is observed, where $Y^n(m)$ is the output of passing $X^n(m)$ through a DMC with transition matrix $P_{Y|X}$ for $m \in \mathcal{M}$, i.e.,

$$P_{Y^n|X^n}(Y^n(m)|X^n(m)) = \prod_{i=1}^n P_{Y|X}(Y_i(m)|X_i(m)). \quad (1)$$

Subsequently, the observed noisy version of the feature vectors are compressed before stored in the database using a deterministic function

$$f^{(n)} : \mathcal{Y}^n \rightarrow \mathcal{L} := \{1, \dots, L\}. \quad (2)$$

For convenience, let $S(m) = f(Y^n(m))$ for all $m \in \mathcal{M}$.

In the identification phase, we are given an index W which is uniformly generated from the set \mathcal{M} and independent of $\{X^n(m), Y^n(m), S(m)\}_{m \in \mathcal{M}}$. The index W is unknown to the database users. Given W , database users observe Z^n , which is the output of passing the feature vector $X^n(W)$ through a DMC with transition matrix $P_{Z|X}$, i.e.,

$$P_{Z^n|X^n}^n(Z^n(W)|X^n(W)) = \prod_{i=1}^n P_{Z|X}(Z_i(W)|X_i(W)). \quad (3)$$

Note that $Z^n - X^n(W) - Y^n(W)$ forms a Markov chain. The user aims to identify W using Z^n and the compressed codebook $\{S(m)\}_{m \in \mathcal{M}}$ using the following deterministic identification function:

$$g^{(n)} : \mathcal{L}^M \times \mathcal{Z}^n \rightarrow \mathcal{M}. \quad (4)$$

Let $\hat{W} := g^{(n)}(S(1), \dots, S(M), Z^n)$ be the estimate of the user. Given the deterministic decoding function $g^{(n)}$, we can define the following disjoint decoding regions

$$\mathcal{D}(S(1), \dots, S(M), W) := \{z^n : g^{(n)}(S(1), \dots, S(M), z^n) = W\}. \quad (5)$$

Finally, in the lossy recovery phase, we need to reproduce the feature vector $X^n(W)$ in a lossy manner using Z^n and $S(\hat{W})$ with a deterministic function

$$h^{(n)} : \mathcal{L} \times \mathcal{Z}^n \rightarrow \hat{\mathcal{X}}^n. \quad (6)$$

Let $\hat{X}^n = h(S(\hat{W}), Z^n)$ be the reproduced feature vector. Define the joint identification-error and excess-distortion probability as follows:

$$\begin{aligned} & P_e^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \\ &:= \Pr \left(\hat{W} \neq W \text{ or } d(X^n(W), \hat{X}^n) > D \right) \end{aligned} \quad (7)$$

$$\begin{aligned} &= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{x^n(1), \dots, x^n(M) \\ y^n(1), \dots, y^n(M) \\ s(1), \dots, s(M)}} \left(P_X^n(x^n(m)) P_{Y|X}^n(y^n(m)|x^n(m)) \prod_{m=1}^M 1\{s(m) = f^{(n)}(y^n(m))\} \right) \\ &\quad \times \sum_{\substack{(z^n, \hat{x}^n): d(x^n(w), \hat{x}^n) > D \text{ or} \\ z^n \notin \mathcal{D}(s(1), \dots, s(M), w)}} P_{Z|X}^n(z^n|x^n(w)) 1\{\hat{x}^n = h^{(n)}(s(w), z^n)\}. \end{aligned} \quad (8)$$

Note that in (8) there are three sources of randomness: i) the randomness of feature vectors $x^n(m) \in \mathcal{X}^n$ for each $m \in \mathcal{M}$ in the enrollment phase; ii) the randomness of $w \in \mathcal{M}$ in identification phase; iii) the randomness $y^n(m) \in \mathcal{Y}^n$ ($m \in \mathcal{M}$) and $z^n(w) \in \mathcal{Z}^n$ due to the two DMCs.

Throughout the paper, we will consider the source distribution being P_X and the two DMCs with transition matrices $P_{Y|X}$ and $P_{Z|X}$. We will use P_{XYZ} to denote $P_X \times P_{Y|X} \times P_{Z|X}$. In all the definitions, we will omit the dependence on distributions $P_X, P_{Y|X}, P_{Z|X}$ for simplicity.

B. Existing Results

First, we present the definition of the rate-distortion region as follows.

Definition 1. A rate-distortion triple (R^i, R^c, D) is said to be ϵ -achievable if there exists a sequence encoding-decoding-reproduction functions $(f^{(n)}, g^{(n)}, h^{(n)})$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M \geq R^i, \quad (9)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log L \leq R^c, \quad (10)$$

$$\limsup_{n \rightarrow \infty} P_e^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq \epsilon. \quad (11)$$

The closure of all ϵ -achievable rate-distortion tuples is called the ϵ -rate-distortion region and denoted as $\mathcal{R}(\epsilon)$.

Let

$$\mathcal{R} = \bigcap_{\epsilon \in (0,1)} \mathcal{R}(\epsilon). \quad (12)$$

In the following, we recall the rate-distortion region by Tuncel and Gündüz [2, Theorem 1]. We remark that their rate-distortion region appears to be identical with \mathcal{R} although it was derived under the average distortion criterion.

Let U be the random variable taking values in the alphabet \mathcal{U} . Define a set of joint distributions on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{U} \times \hat{\mathcal{X}}$ as

$$\mathcal{P}^* := \left\{ Q_{XYZU\hat{X}} : |\mathcal{U}| \leq |\mathcal{Y}| + 2, Z - X - Y - U, Q_X = P_X, Q_{Y|X} = P_{Y|X}, Q_{Z|X} = P_{Z|X}, \right. \\ \left. \hat{X} = \phi(U, Z) \text{ for some } \phi : \mathcal{U} \times \mathcal{Z} \rightarrow \hat{\mathcal{X}} \right\}. \quad (13)$$

Given $Q_{XYZU\hat{X}}$, let

$$\mathcal{R}(Q_{XYZU\hat{X}}) = \left\{ (R^i, R^c, D) : R^i \leq I(Q_U, Q_{Z|U}), R^c - R^i \geq I(Q_{U|Z}, Q_{Y|UZ}|Q_Z), D \geq \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})] \right\}, \quad (14)$$

and let

$$\mathcal{R}^* := \bigcup_{Q_{XYZU\hat{X}} \in \mathcal{P}^*} \mathcal{R}(Q_{XYZU\hat{X}}). \quad (15)$$

Theorem 1. The rate-distortion region \mathcal{R} satisfies

$$\mathcal{R} = \mathcal{R}^*. \quad (16)$$

III. MAIN RESULTS: EXPONENTIAL STRONG CONVERSE THEOREM

A. Preliminaries

In this subsection, we make necessary definitions and present a key lemma to ensure the smooth presentation of exponential strong converse theorem in Section III-B.

Let

$$\mathcal{Q} := \left\{ Q_{XYZU\hat{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{U} \times \hat{\mathcal{X}}) : |\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}||\mathcal{Z}||\hat{\mathcal{X}}| \right\}. \quad (17)$$

Given $(\alpha, \mu, \beta, \theta) \in [0, 1]^3 \times \mathbb{R}_+$ and a distribution $Q_{XYZU\hat{X}} \in \mathcal{Q}$, define the linear combination of the likelihood ratios

$$\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u) := \bar{\alpha} \left(\log \frac{Q_Y(y)}{P_Y(y)} + \log \frac{Q_{Z|YU}(z|y, u)}{P_{Z|Y}(x|y)} + \log \frac{Q_{X|YZU}(x|y, z, u)}{P_{X|YZ}(x|y, z)} \right. \\ \left. + \log \frac{Q_{XY|ZU\hat{X}}(\hat{x}|x, y, z, u)}{Q_{XY|ZU}(x, y|z, u)} \right) + \alpha \left(\bar{\mu} \bar{\beta} \log \frac{Q_{YZ|U}(y, z|u)}{P_{YZ}(y, z)} + \bar{\mu} \log \frac{Q_Z(z)}{Q_{Z|U}(z|u)} + \mu d(x_i, \hat{x}_i) \right). \quad (18)$$

Also define the negative cumulant generating functions as

$$\Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}) := -\log \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\exp \left(-\theta \omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right) \right], \quad (19)$$

$$\Omega^{(\alpha, \mu, \beta, \theta)} := \min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}). \quad (20)$$

Finally, we define the large-deviation rate functions

$$F^{(\alpha, \mu, \beta, \theta)}(R^i, R^c, D) := \frac{\Omega^{(\alpha, \mu, \beta, \theta)} - \theta \alpha (\bar{\mu}(\bar{\beta} R^c - R^i) + \mu D)}{1 + 5\theta \bar{\alpha} + \theta \alpha \bar{\mu}(3 - \beta)}, \quad (21)$$

$$F(R^i, R^c, D) := \sup_{(\alpha, \mu, \beta, \theta) \in [0, 1]^3 \times \mathbb{R}_+} F^{(\alpha, \mu, \beta, \theta)}(R^i, R^c, D). \quad (22)$$

Let

$$\rho = \sup_{Q_{XY ZU \hat{X}} \in \mathcal{Q}} \text{Var}_{Q_{XY ZU \hat{X}}} \left[\omega_{Q_{XY ZU \hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right]. \quad (23)$$

We then have the following conclusions on the quantities defined in (22).

Lemma 2. *The following conclusions hold.*

- i) *For any $\delta > 0$, there exists a positive number ν such that for every $\tau \in (0, \nu]$, if $(R^c + \tau, R^i - \tau, D + \tau) \notin \mathcal{R}$, then we obtain*

$$F(R^i, R^c, D) \geq \frac{\rho}{2} \phi^2 \left(\frac{\tau^{3+\delta}}{\rho} \right) > 0, \quad (24)$$

where ϕ is the inverse function of $\psi(a) := 2a + 8a^2$ for $a > 0$. Hence, if $(R^i, R^c, D) \notin \mathcal{R}$, then

$$F(R^i, R^c, D) > 0; \quad (25)$$

- ii) *If $(R^i, R^c, D) \in \mathcal{R}$, then*

$$F(R^i, R^c, D) = 0. \quad (26)$$

The proof of Lemma 2 is similar to that of [20, Property 4] and is given in Appendix E. We remark that Lemma 2, especially conclusion i), plays an central role in claiming the exponential strong converse theorem for the content identification problem with lossy recovery. As we will see shortly in Theorem 3, $F(R^i, R^c, D)$ in (22) is a lower bound on the exponent of the probability of correct decoding.

B. Exponential Strong Converse

Theorem 3. *For any encoding-decoding functions $(f^{(n)}, g^{(n)})$ such that*

$$\frac{1}{n} \log L \leq R^c, \quad (27)$$

$$\frac{1}{n} \log M \geq R^i, \quad (28)$$

given any deterministic function $h^{(n)}$ and any distortion level D , we have

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq 7 \exp(-nF(R^i, R^c, D)). \quad (29)$$

The proof of Theorem 3 is given in Section IV. In the proof, we adapt the information spectrum method proposed by Oohama [18]–[20] to first establish a non-asymptotic upper bound on the probability of correct decoding. Invoking the upper bound (cf. Lemma 10) and applying Cramér's theorem on large deviations, we can further upper bound the probability of correct decoding. Subsequently, we proceed in a similar manner as [19], [20] to obtain the desired lower bound.

Second, we believe that both the image size characterization [25] and the perturbation approach [26] probably cannot lead to the strong converse theorem for the content identification problem with lossy recovery. The major difficulty lies in the fact that decoder needs to use the whole codebook $\mathcal{C} = \{S(1), \dots, S(M)\}$ and Z^n to decode. Recall that $S(m) = f^{(n)}(Y^n(m))$ for $m = 1, \dots, M$.

Invoking Lemma 2 and Theorem 3, we conclude that the exponent in the right hand side of (29) is strictly positive if the rate pairs are outside the rate-distortion region. Hence, we obtain the following exponential strong converse theorem.

Theorem 4. *For any sequence of encoding-decoding-reproduction functions $(f^{(n)}, g^{(n)}, h^{(n)})$ such that*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log L \leq R^c, \quad (30)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M \geq R^i, \quad (31)$$

given a distortion level D , we have that if $(R^i, R^c, D) \notin \mathcal{R}$ (recall Theorem 1), then the probability of correct decoding vanishes to zero exponentially fast as n goes to infinity.

Invoking Theorem 4, we conclude that the ϵ -rate distortion region satisfies $\mathcal{R}(\epsilon) = \mathcal{R}^*$. Adopting the one-shot technique introduced in [27], we can also establish a non-asymptotic achievability bound. Applying the Berry-Esseen theorem to the achievability bound and analyzing the bound in Lemma 3, we can conclude that the second-order coding region is in the order of $\Omega(1/n^{1/2}) \cap O(1/n^{1/6})$. Nailing down the exact characterization of the second-order asymptotics is left as future work.

C. Upper Bound on Joint Error and Excess-distortion Probability

Definition 2. A non-negative number E is said to be an (R^i, R^c, D) -achievable joint error and excess-distortion exponent if there exists a sequence of encoding-decoding-reproduction functions $(f^{(n)}, g^{(n)}, h^{(n)})$ such that (30), (31) hold and

$$\limsup_{n \rightarrow \infty} -\frac{\log P_e^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D)}{n} \leq E. \quad (32)$$

The supremum of all (R^i, R^c, D) -achievable error exponent is called the optimal error exponent and denoted as $E(R^i, R^c, D)$.

Recall that \mathcal{R} (Definition 1) is the rate-distortion region with respect to $P_X, P_{Y|X}, P_{Z|X}$. For any $Q_X, Q_{Y|X}, Q_{Z|X}$, let $\mathcal{R}(Q_X, Q_{Y|X}, Q_{Z|X})$ be the rate-distortion region with respect to $Q_X, Q_{Y|X}, Q_{Z|X}$. Invoking Lemma 2, Theorem 3 and applying Marton's change-of-measure technique [28], we derive an upper bound on $E(R^i, R^c, D)$.

Theorem 5. The optimal joint error and excess-distortion exponent function satisfies

$$E(R^i, R^c, D) \leq \inf_{\substack{Q_{XYZ}: Z-X-Y \\ (R^i, R^c, D) \notin \mathcal{R}(Q_X, Q_{Y|X}, Q_{Z|X})}} D(Q_{XYZ} \| P_{XYZ}). \quad (33)$$

Our main results for content identification with lossy recovery (Theorems 4 and 5) can be specialized to the biometrical identification problem [6], the content identification problem [5] and the Wyner-Ziv source coding problem [24] since all these three problems are special cases of the content identification problem with lossy recovery as argued in [2].

D. Extensions for the Biometrical Identification Problem

In this subsection, we present several extensions for the biometrical identification problem [6]. The capacity (the maximum rate) of the biometrical identification problem was characterized by Willems, Kalker, Goseling and Linnartz in [6]. Further, the exponential strong converse theorem for the biometrical identification problem has been established in [9, Theorem 2].

Compared with the content identification problem with lossy recovery, there is no compression (no $f^{(n)}$) and no lossy recovery phase (no $h^{(n)}$) in the biometrical identification problem. Thus, $S(m) = Y^n(m)$ for each $m \in \mathcal{M}$. Hence, the error probability of the biometrical identification problem is

$$P_e^{(n)}(g^{(n)}) := \Pr(\hat{W} \neq W). \quad (34)$$

Let C_{bio} be the capacity of the biometrical identification problem. Then, it can be verified that

$$C_{\text{bio}} = \sup\{R^i : (R^i, \log |\mathcal{Y}|, d^+) \in \mathcal{R}\} \quad (35)$$

$$= I(P_Y, P_{Z|Y}). \quad (36)$$

Define the exponents

$$\underline{E}_{\text{bio}}(R^i) := \sup_{\lambda > 0} \frac{1}{1 + \lambda} \left(\lambda R^i - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right] \right), \quad (37)$$

$$\overline{E}_{\text{bio}}(R^i) := \sup_{\lambda > 0} \lambda R^i - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right]. \quad (38)$$

Theorem 6. For any decoding function $g^{(n)}$, we have that

$$P_c^{(n)}(g^{(n)}) \leq 2 \exp(-n \underline{E}_{\text{bio}}(R^i)). \quad (39)$$

Further, there exists a decoding function $g^{(n)}$ such that

$$P_c^{(n)}(g^{(n)}) \geq \frac{1}{2} \exp(-n \overline{E}_{\text{bio}}(R^i)). \quad (40)$$

It is easy to verify that $\underline{E}_{\text{bio}}(R^i) > 0$ if $R^i > C_{\text{bio}} = I(P_Y, P_{Z|Y})$ and $\overline{E}_{\text{bio}} = 0$ if $R^i \leq C_{\text{bio}}$. Hence, the exponential strong converse theorem follows as a simple corollary. Although we cannot establish a tight strong converse exponent, in the following, we present tight results on the moderate deviations constant. Let $P_c^*(n, R^i)$ be the maximum probability of correct decoding when the number of items to be distinguished M satisfies that $\log M \geq nR^i$. Let

$$V := \text{Var} \left[\log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right]. \quad (41)$$

Throughout this section, we assume that $V > 0$.

Theorem 7. Consider any sequence of positive numbers $\{\xi_n\}_{n=1}^\infty$ such that $\xi_n \rightarrow 0$ and $\sqrt{n}\xi_n \rightarrow \infty$ as $n \rightarrow \infty$. When the rate R^i approaches capacity C_{bio} from above, the probability of correct decoding scales as

$$\lim_{n \rightarrow \infty} -\frac{\log P_c^*(n, C_{\text{bio}} + \xi_n)}{n\xi_n^2} = \frac{1}{2V}. \quad (42)$$

Similarly, when the rate R^i approaches capacity C_{bio} from below, the probability of correct decoding scales as

$$\lim_{n \rightarrow \infty} -\frac{\log(1 - P_c^*(n, C_{\text{bio}} - \xi_n))}{n\xi_n^2} = \frac{1}{2V}. \quad (43)$$

The result in (42) implies that even if the rate R^i approaches the capacity from with speed ξ_n , the probability of correct decoding still vanishes to zero (subexponentially fast). Similarly, (43) implies that if the rate R^i approaches the capacity from below with speed ξ_n , then the probability of error vanishes to zero (subexponentially fast).

We remark that the study of moderate deviations for DMCs was done by Altuğ and Wagner [29] and also by Polyanskiy and Verdú [30]. For certain classes of quantum channels, moderate deviations analysis (above and below capacity) was done by Chubb, Tan, and Tomamichel [31]. For other works on moderate deviations, see [32]–[34].

In the following, we also present the tradeoff between the number of items to be distinguished and the error probability when it is non-vanishing. Let $M^*(n, \epsilon)$ be the maximum number of items to be distinguished such that the error probability satisfies $P_e^{(n)}(g^{(n)}) \leq \epsilon$ for some decoding function $g^{(n)}$. The second-order coding rate for the biometrical identification problem is defined as

$$L^*(\epsilon) := \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} (\log M^*(n, \epsilon) - nI(P_Y, P_{Z|Y})). \quad (44)$$

Theorem 8. For any $\epsilon \in (0, 1)$, the second-order coding rate for the biometrical identification problem satisfies

$$L^*(\epsilon) = \sqrt{V}\Phi^{-1}(\epsilon). \quad (45)$$

The result in Theorem 8 implies that if we allow a non-vanishing error probability, then the rate $R^i(n, \epsilon) := \frac{1}{n} \log M^*(n, \epsilon)$ approaches capacity C_{bio} with speed $L^*(\epsilon)/\sqrt{n}$.

We remark that the second-order asymptotics date back to Strassen [35] and revived by Hayashi [36] and by Polyanskiy, Poor and Verdú [37]. For a summary of works on second-order asymptotics, see [38].

The proofs of Theorems 6, 7 and 8 are given in Appendix F.

IV. PROOF OF THEOREM 3

In this section, we present the proof of Theorem 3.

A. Preliminaries

To ensure the smooth presentation of the non-asymptotic converse bound in Lemmas 9, 10, in this subsection, we present necessary definitions. Given an encoding function $f^{(n)}$ and any $m \in \mathcal{M}$, let

$$P_{S|Y^n}(s(m)|y^n(m)) := 1\{s(m) = f^{(n)}(y^n(m))\}. \quad (46)$$

Given a deterministic function $h^{(n)}$ and any $w \in \mathcal{M}$, let

$$P_{\hat{X}^n|SZ^n}(\hat{x}^n|s(w), z^n) := 1\{\hat{x}^n = h^{(n)}(s(w), z^n)\}. \quad (47)$$

For simplicity, we use \mathbf{x} to denote x^n , \mathbf{x}^M to denote $(x^n(1), \dots, x^n(M))$, s^M to denote $s(1), \dots, s(M)$. In a similar manner, we have $\mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}, \mathbf{y}^M$ and the corresponding random vectors $\mathbf{X}, \mathbf{X}^M, \mathbf{Y}, \mathbf{Y}^M, S^M, \mathbf{Z}, \hat{\mathbf{X}}$. For simplicity, let $\mathbf{t} := (\mathbf{x}^M, \mathbf{y}^M, s^M, z^n, \hat{x}^n)$, $\mathbf{T} := (\mathbf{X}^M, \mathbf{Y}^M, S^M, \mathbf{Z}, \hat{\mathbf{X}})$ and $\mathcal{T} = (\mathcal{X}^{Mn}, \mathcal{Y}^{Mn}, \mathcal{L}^M, \mathcal{Z}^n, \hat{\mathcal{X}}^n)$. Then let $P_{W\mathbf{T}}$ be the joint distribution of $(W, \mathbf{X}^M, \mathbf{Y}^M, S^M, \mathbf{Z}, \hat{\mathbf{X}})$, induced by $P_W, P_X^n, P_{Y|X}^n, P_{Z|X}^n, P_{Y|S^n}, P_{\hat{X}^n|SZ^n}$, i.e.,

$$P_{W\mathbf{T}}(w, \mathbf{t}) = \frac{1}{M} \left(\prod_{m=1}^M P_X^n(x^n(m)) P_{Y|X}^n(Y^n(m)|x^n(m)) P_{S|Y^n}(s(m)|y^n(m)) \right) P_{Z|X}^n(z^n|x^n(w)) P_{\hat{X}^n|SZ^n}(\hat{x}^n|s(w), z^n). \quad (48)$$

Further, in this section, whenever we use \mathbb{E} , we mean the expectation over $P_{W\mathbf{T}}$ unless otherwise stated. Note that for each $w \in \mathcal{M}$, the joint distribution of $(\mathbf{X}(w), \mathbf{Y}(w), S(w), \mathbf{Z}, \hat{\mathbf{X}})$ is the same and let the joint distribution be denoted as $P_{\mathbf{X}\mathbf{Y}SZ\hat{\mathbf{X}}}$. In the following, all the distributions starting with P are induced by $P_{\mathbf{X}\mathbf{Y}SZ\hat{\mathbf{X}}}$.

Let $Q_{Y^n}, Q_{X^n|SY^nZ^n}, Q_{X^nY^n|SZ^n\hat{X}^n}, Q_{Y^nZ^n|S}, Q_{Z^n}$ be arbitrary distributions. Given $w \in \mathcal{M}$ and any $\eta > 0$, define the following sets:

$$\mathcal{A}_1(w) := \left\{ \mathbf{t} : \frac{1}{n} \log \frac{P_{Y^n}(y^n(w))}{Q_{Y^n}(y^n(w))} \geq -\eta \right\}, \quad (49)$$

$$\mathcal{A}_2(w) := \left\{ \mathbf{t} : \frac{1}{n} \log \frac{P_{Z^n|Y^n}(z^n|y^n(w))}{Q_{Z^n|SY^n}(z^n|s(w), y^n(w))} \geq -\eta \right\}, \quad (50)$$

$$\mathcal{A}_3(w) := \left\{ \mathbf{t} : \frac{1}{n} \log \frac{P_{X^n|Y^nZ^n}(x^n(w)|y^n(w), z^n)}{Q_{X^n|SY^nZ^n}(y^n(w)|s(w), x^n(w), z^n)} \geq -\eta \right\}, \quad (51)$$

$$\mathcal{A}_4(w) := \left\{ \mathbf{t} : \frac{1}{n} \log \frac{P_{X^nY^n|SZ^n}(x^n(w), y^n(w)|s(w), z^n)}{Q_{X^nY^n|SZ^n\hat{X}^n}(x^n(w), y^n(w)|s(w), z^n, \hat{x}^n)} \geq -\eta \right\}, \quad (52)$$

$$\mathcal{A}_5(w) := \left\{ \mathbf{t} : R^c \geq \frac{1}{n} \log \frac{Q_{Y^nZ^n|S}(y^n(w)z^n|s(w))}{P_{Y^nZ^n}(y^n(w), z^n)} - \eta \right\}, \quad (53)$$

$$\mathcal{A}_6(w) := \left\{ \mathbf{t} : R^i \leq \frac{1}{n} \log \frac{P_{Z^n|S}(z^n|s(w))}{Q_{Z^n}(z^n)} + \eta \right\}, \quad (54)$$

$$\mathcal{A}_7(w) := \left\{ \mathbf{t} : d(x^n(w), \hat{x}^n) \leq D \right\}. \quad (55)$$

Choose $U_i(W) = (S(W), Y^{i-1}(W), Z_{i+1}^n)$. Then it can be verified that $Z_i - X_i(W) - Y_i(W) - U_i(W)$ and $(X_i(W), Y_i(W)) - (U_i(W), Z_i) - \hat{X}_i$ form two Markov chains under the joint distribution $P_{W\mathbf{T}}$ (recall (48)). Further, let $V_i(W) = (S(W), Z_{i+1}^n)$.

For $i = 1, \dots, n$, let $Q_{X_iY_iZ_iU_i\hat{X}_i}$ be any generic distributions and let $Q_{Y_i}, Q_{Z_i}, Q_{X_i|Y_iZ_iU_i}, Q_{X_iY_i|Z_iU_i\hat{X}_i}, Q_{Y_iZ_i|U_i}$ be induced by $Q_{X_iY_iZ_iU_i\hat{X}_i}$. Paralleling (49) to (54), given any $\eta > 0$, we define the following sets:

$$\mathcal{B}_1(w) := \left\{ \mathbf{t} : 0 \geq \frac{1}{n} \sum_{i=1}^n \log \frac{Q_{Y_i}(y_i(w))}{P_{Y_i}(y_i(w))} - \eta \right\}, \quad (56)$$

$$\mathcal{B}_2(w) := \left\{ \mathbf{t} : 0 \geq \frac{1}{n} \sum_{i=1}^n \frac{Q_{Z_i|Y_iU_i}(x_i(w)|z_i, u_i(w))}{P_{Z_i|Y_i}(z_i|y_i(w))} - \eta \right\}, \quad (57)$$

$$\mathcal{B}_3(w) := \left\{ \mathbf{t} : 0 \geq \frac{1}{n} \sum_{i=1}^n \log \frac{Q_{X_i|Y_iZ_iU_i}(x_i(w)|y_i(w), z_i, u_i(w))}{P_{X_i|Y_iZ_i}(x_i(w)|y_i(w), z_i)} - \eta \right\}, \quad (58)$$

$$\mathcal{B}_4(w) := \left\{ \mathbf{t} : 0 \geq \frac{1}{n} \sum_{i=1}^n \log \frac{Q_{X_iY_i|Z_iU_i\hat{X}_i}(x_i(w), y_i(w)|z_i, u_i(w), \hat{x}_i)}{P_{X_iY_i|Z_iU_i}(x_i(w), y_i(w)|z_i, u_i(w))} - \eta \right\}, \quad (59)$$

$$\mathcal{B}_5(w) := \left\{ \mathbf{t} : R^c - R^i \geq \frac{1}{n} \sum_{i=1}^n \log \left(\frac{Q_{Y_iZ_i|U_i}(y_i(w), z_i|u_i(w))}{P_{Y_iZ_i}(y_i(w), z_i)} \frac{Q_{Z_i}(z_i)}{P_{Z|V_i}(z_i|v_i(w))} \right) - 3\eta \right\}, \quad (60)$$

$$\mathcal{B}_6(w) := \left\{ \mathbf{t} : R^i \leq \frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z_i|V_i}(z_i|v_i(w))}{Q_{Z_i}(z_i)} + \eta \right\}, \quad (61)$$

$$\mathcal{B}_7(w) := \left\{ \mathbf{t} : D \geq \frac{1}{n} \sum_{i=1}^n \log e^{d(x_i(w), \hat{x}_i)} \right\}. \quad (62)$$

B. Proof of Theorem 3

Invoking (8) and (48), we define the probability of correct decoding as

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) := 1 - P_e^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \quad (63)$$

$$= \sum_{w=1}^M \sum_{\substack{t \in \mathcal{T}: \mathbf{z} \in \mathcal{D}(s^M, w) \\ d(\mathbf{x}(w), \hat{\mathbf{x}}) \leq D}} P_{W\mathbf{T}}(w, \mathbf{t}). \quad (64)$$

We first present a non-asymptotic upper bound on $P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D)$, i.e., a non-asymptotic converse bound for the problem.

Lemma 9. For any encoding-decoding functions $(f^{(n)}, g^{(n)})$ such that

$$\frac{1}{n} \log L \leq R^c, \quad (65)$$

$$\frac{1}{n} \log M \geq R^i, \quad (66)$$

given any deterministic function $h^{(n)}$ and any distortion level D , we have

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq P_{WT} \left\{ \bigcap_{i=1}^7 \mathcal{A}_i(W) \right\} + 6e^{-n\eta}. \quad (67)$$

The proof of Lemma 9 is given in Appendix A.

A few other remarks are in order. First, in the proof of Lemma 9, we define seven sets for each $w \in \mathcal{M}$ in (49)–(55). Equipped with these definitions, we obtain the upper bound in (67) where the probability of correct decoding of W depends on $S(W), X^n(W), Y^n(W), Z^n, \hat{X}^n$.

Second, in (67), $\mathcal{A}_5(w)$ corresponds to the enrollment phase, $\mathcal{A}_6(w)$ corresponds to the identification phase and $\mathcal{A}_7(w)$ corresponds to the lossy recovery phase. Further, $\mathcal{A}_1(w)$ to $\mathcal{A}_4(w)$ are the auxiliary sets whose roles will be clear in subsequent analyses.

Third, the definitions of \mathcal{Q} (cf. (17)) and $\{\mathcal{A}_i(w)\}_{i=1}^7$ are crucial. Note that $\mathcal{A}_1(w)$ to $\mathcal{A}_4(w)$ appear in Lemma 9. They appear due to the different Markov conditions in the definitions of \mathcal{P}^* in (13) and \mathcal{Q} in (17). This is also closely related with the proof of Lemma 2 in Appendix E. Hence, there is a subtle interplay between Lemmas 9 and 2. This tension also appears in [20]. However, we need to adapt [20] to the setting here of the content identification problem with lossy recovery carefully since these two problems are significantly different.

Invoking Lemma 9 and choosing the distributions $Q_{Y^n}, Q_{X^n|SY^nZ^n}, Q_{X^nY^n|SZ^n\hat{X}^n}, Q_{Y^nZ^n|S}, Q_{Z^n}$ properly, we obtain the following lemma.

Lemma 10. *Given the conditions in Lemma 9, we have*

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq P_{WT} \left\{ \bigcap_{i=1}^7 \mathcal{B}_i(W) \right\} + 6e^{-n\eta}. \quad (68)$$

The proof of Lemma 10 is given in Appendix B.

A few remarks are as follow. First, in order to introduce $U_i(W)$ into the bound in Lemma 10, we make use of the Markov chain $(X_i(W), Y_i(W)) - (S(W), Y^{i-1}(W), Z_i^n) - (X^{i-1}(W), Z^{i-1})$ which can be established similarly as [20, Lemma 2].

Second, note that in [2], Tuncel and Gündüz chose the auxiliary random variable as $U_i(W) = (S(W), Z^{i-1}, Z_{i+1}^n)$. If we choose $U_i(W)$ as in [2], we can not obtain Lemma 10. Further, note that here we use both $U_i(W)$ and $V_i(W)$. This idea is also present in Oohama's work [19] on degraded broadcast channel. In the subsequent analysis in Lemma 12, we will eliminate $V_i(W)$ by using Hölder's inequality.

In the following, for simplicity, we will use Q_i to denote $Q_{X_iY_iZ_iU_i\hat{X}_i}$ and use P_i to denote $P_{X_iY_iZ_iU_iV_i\hat{X}_i}$. Let $(\alpha, \mu, \beta, \lambda) \in [0, 1]^3 \times \mathbb{R}_+$ be given. Then we need the following definitions to further upper bound the right hand side in Lemma 10. Let

$$\begin{aligned} f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i) &:= \frac{Q_{Y_i}^{\bar{\alpha}}(y_i)}{P_{Y_i}^{\bar{\alpha}}(y_i)} \frac{Q_{Z_i|Y_iU_i}^{\bar{\alpha}}(z_i | y_i, u_i)}{P_{Z_i|Y_i}^{\bar{\alpha}}(z_i | y_i)} \frac{Q_{X_i|Y_iZ_iU_i}^{\bar{\alpha}}(x_i | y_i, z_i, u_i)}{P_{X_i|Y_iZ_i}^{\bar{\alpha}}(x_i | y_i, z_i)} \\ &\times \frac{Q_{X_iY_i|Z_iU_i\hat{X}_i}^{\bar{\alpha}}(x_i, y_i | z_i, u_i(w), \hat{x}_i)}{P_{X_iY_i|Z_iU_i}^{\bar{\alpha}}(x_i, y_i | z_i, u_i)} \frac{Q_{Y_iZ_i|U_i}^{\alpha\bar{\mu}\bar{\beta}}(y_i, z_i | u_i)}{P_{Y_iZ_i}^{\alpha\bar{\mu}\bar{\beta}}(y_i, z_i)} \frac{Q_{Z_i}^{\alpha\bar{\mu}}(z_i)}{P_{Z_i|V_i}^{\alpha\bar{\mu}}(z_i | v_i)} e^{\alpha\mu d(x_i, \hat{x}_i)}. \end{aligned} \quad (69)$$

Then, let

$$\Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) := -\log \mathbb{E} \left[\exp \left(-\lambda \sum_{i=1}^n \log f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(X_i(W), Y_i(W), Z_i, \hat{X}_i | U_i(W), V_i(W)) \right) \right]. \quad (70)$$

Invoking Cramér's bound on large deviations (cf. Lemma 13), we obtain the following lemma.

Lemma 11. *For any $(\alpha, \mu, \beta, \lambda) \in [0, 1]^3 \times \mathbb{R}_+$, given the conditions in Lemma 9, we have*

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq 7 \exp \left(-n \frac{\frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) - \lambda \alpha \bar{\mu} (\bar{\beta} R^c - R^i) - \lambda \alpha \mu D}{1 + \lambda (4\bar{\alpha} + \alpha \bar{\mu} (3 - 2\beta))} \right). \quad (71)$$

The proof of Lemma 11 is given in Appendix C.

Let

$$\underline{\Omega}^{(\alpha, \mu, \beta, \lambda)} := \inf_{n \geq 1} \sup_{\{Q_i\}_{i=1}^n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n). \quad (72)$$

Define

$$\theta := \frac{\lambda}{1 - \lambda \bar{\alpha} - \lambda \alpha \bar{\mu} \beta}. \quad (73)$$

Hence, we have

$$\lambda = \frac{\theta}{1 + \theta\bar{\alpha} + \theta\alpha\bar{\mu}\beta}. \quad (74)$$

The next lemma is essential in the proof.

Lemma 12. *For any $(\alpha, \mu, \beta, \lambda) \in [0, 1]^3 \times \mathbb{R}_+$ such that $\lambda \in (0, \frac{1}{\bar{\alpha} + \alpha\bar{\mu}\beta})$, we have*

$$\underline{\Omega}^{(\alpha, \mu, \beta, \lambda)} \geq \frac{\Omega^{(\alpha, \mu, \beta, \theta)}}{1 + \theta\bar{\alpha} + \theta\alpha\bar{\mu}\beta}. \quad (75)$$

The proof of Lemma 12 is similar to that of [20, Proposition 2] and given in Appendix D. In the proof of Lemma 12, we first remove W in the expression of $\Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n)$. Subsequently, by adopting ideas from [20] and [19] and properly choosing distributions $Q_{X_i Y_i Z_i U_i \hat{X}_i}$ via the recursive method, we can establish Lemma 12.

Invoking Lemmas 11 and 12, we conclude that

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \quad (76)$$

$$\leq 7 \exp \left(-n \frac{\underline{\Omega}^{(\alpha, \mu, \beta, \lambda)} - \lambda\alpha\bar{\mu}(\bar{\beta}R^c - R^i) - \lambda\alpha\mu D}{1 + \lambda(4\bar{\alpha} + \alpha\bar{\mu}(3 - 2\beta))} \right) \quad (77)$$

$$\leq 7 \exp \left(-n \frac{\frac{\Omega^{(\alpha, \mu, \beta, \theta)}}{1 + \theta\bar{\alpha} + \theta\alpha\bar{\mu}\beta} - \frac{\theta\alpha\bar{\mu}(\bar{\beta}R^c - R^i) + \theta\alpha\mu D}{1 + \theta\bar{\alpha} + \theta\alpha\bar{\mu}\beta}}{1 + \frac{\theta(4\bar{\alpha} + \alpha\bar{\mu}(3 - 2\beta))}{1 + \theta\bar{\alpha} + \theta\alpha\bar{\mu}\beta}}} \right) \quad (78)$$

$$= 7 \exp \left(-n \frac{\Omega^{(\alpha, \mu, \beta, \theta)} - \theta\alpha\bar{\mu}(\bar{\beta}R^c - R^i) - \theta\alpha\mu D}{1 + 5\theta\bar{\alpha} + \theta\alpha\bar{\mu}(3 - \beta)} \right) \quad (79)$$

$$= 7 \exp(-nF(R^i, R^c, D)), \quad (80)$$

where (80) follows from the definition of $F(R^i, R^c, D)$ in (22). The proof of Theorem 3 is now complete.

V. CONCLUSION

In this paper, we derived a non-asymptotic converse bound for content identification problem with lossy recovery. Invoking the non-asymptotic bound, we established the exponential strong converse theorem. As a corollary of our main results, we derived an upper bound on the optimal exponent of the joint error and excess-distortion. Our main results can be specialized to the biometrical identification problem [6] and the content identification problem [5].

There are several avenues for future research. First, note that in Theorem 3, we present only a non-asymptotic exponential type upper bound on the probability of correct decoding. Although this is sufficient for us to claim the exponential strong converse theorem (cf. Theorem 4) by invoking Lemma 2, it is worth deriving the exact exponent for the probability of correct decoding. The ideas involved in characterizing the exact strong converse exponent in [39], [40] and the one-shot techniques in [27], [41] might be useful.

Second, after Theorem 4, we remark the second-order coding rates are in the order of $O(1/n^{1/6}) \cap \Omega(1/\sqrt{n})$. In the future, one may be interested in nailing down the exact order. For this line of research, one may borrow ideas from [42]–[44].

Third, in this paper, we only considered the discrete memory sources and discrete memoryless channels. In future, one may consider Gaussian memoryless sources, the additive Gaussian white noise channel, and the quadratic distortion measure. For the special case of biometrical identification problem, the capacity for Gaussian case was derived in [45]. However, for Gaussian case of content identification with lossy recovery, one has to first calculate the rate-distortion region (cf. Theorem 1). To do so, it is necessary to check whether Gaussian test channels are first-order optimal by referring to [46] and [47]. The strong converse theorem for Gaussian case may be inspired by works of Fong and Tan in [48] and [49].

APPENDIX

A. Proof of Lemma 9

Recall that $\mathbf{x} = x^n, \mathbf{y} = y^n, \mathbf{z} = z^n, \mathbf{t} = (\mathbf{x}^M, \mathbf{y}^M, s^M, \mathbf{z}, \hat{\mathbf{x}})$ and we will drop the subscript of distributions when necessary. Recall the definition of the distribution $P_{W\mathbf{T}}$ in (48) and the definitions of $\{\mathcal{A}_i(w)\}_{i=1}^7$ in (49) to (54). Invoking (64) and noting that $d(\mathbf{x}(w), \hat{\mathbf{x}}) \leq D$ is equivalent to $t \in \mathcal{A}_7(w)$, we have the probability of correct decoding as follows:

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) = \sum_{w=1}^M \left(\sum_{\substack{\mathbf{t} \in \left(\left(\bigcap_{i=1}^6 \mathcal{A}_i(w) \right) \cap \mathcal{A}_7(w) \right): \\ z^n \in \mathcal{D}(s^M, w)}} P_{W\mathbf{T}}(w, \mathbf{t}) + \sum_{\substack{\mathbf{t} \in \left(\left(\bigcup_{i=1}^6 \mathcal{A}_i(w) \right) \cap \mathcal{A}_7(w) \right): \\ z^n \in \mathcal{D}(s^M, w)}} P_{W\mathbf{T}}(w, \mathbf{t}) \right). \quad (81)$$

Ignoring the constraint that $z^n \in \mathcal{D}(s^M, w)$, we can upper bound the first term in (81) by

$$\Delta_1 = \sum_{w=1}^M \frac{1}{M} \sum_{\mathbf{t} \in \bigcap_{i=1}^7 \mathcal{A}_i(w)} P(\mathbf{t}) = P_{W\mathbf{T}} \left\{ \bigcap_{i=1}^7 \mathcal{A}_i(W) \right\}. \quad (82)$$

For $i = 2, \dots, 7$, let

$$\Delta_i := \sum_{w=1}^M \sum_{\substack{\mathbf{t} \in \mathcal{A}_{i-1}^c(w) \cap \mathcal{A}_7(w): \\ z^n \in \mathcal{D}(s^M, w)}} P_{W\mathbf{T}}(w, \mathbf{t}). \quad (83)$$

Then, the second term in (81) is no larger than $\sum_{i=2}^7 \Delta_i$ by the union bound.

For simplicity, let $\mathbf{t}(w) := (\mathbf{x}(w), \mathbf{y}(w), s(w), \mathbf{z}, \hat{\mathbf{x}})$. Invoking (49), in a similar manner as the proof of [20, Lemma 12], we obtain that

$$\Delta_2 = \sum_{w=1}^M \sum_{\substack{\mathbf{t} \in \mathcal{A}_1^c(w) \cap \mathcal{A}_7(w): \\ \mathbf{z} \in \mathcal{D}(s^M, w)}} P(\mathbf{t}) \quad (84)$$

$$\leq \sum_{w=1}^M \sum_{\mathbf{t} \in \mathcal{A}_1^c(w)} P(\mathbf{t}) \quad (85)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{\mathbf{t}(w): \\ P(\mathbf{y}(w)) \leq e^{-n\eta} Q(\mathbf{y}(w))}} P(\mathbf{t}(w)) \quad (86)$$

$$\leq \sum_{w=1}^M \frac{1}{M} \sum_{\mathbf{y}(w): P(\mathbf{y}(w)) \leq e^{-n\eta} Q(\mathbf{y}(w))} P(\mathbf{y}(w)) \quad (87)$$

$$\leq e^{-n\eta} \sum_{w=1}^M \frac{1}{M} \sum_{\mathbf{y}(w)} Q(\mathbf{y}(w)) \quad (88)$$

$$\leq e^{-n\eta}, \quad (89)$$

where (85) follows from ignoring the constraints that $(\mathbf{x}^M, \mathbf{y}^M, s^M, \mathbf{z}, \hat{\mathbf{x}}) \in \mathcal{A}_7(w)$ and $z^n \in \mathcal{D}(s^M, w)$; (86) follows from the definition of $\mathcal{A}_1(w)$ in (49) and the fact that for each w ,

$$\sum_{m \neq w, m \in \mathcal{M}} \sum_{\mathbf{x}(m), \mathbf{y}(m), s(m)} P(\mathbf{t}) = P(\mathbf{t}(w)), \quad (90)$$

and (87) follows since $\sum_{\mathbf{x}(w), s(w), \mathbf{z}, \hat{\mathbf{x}}} P(\mathbf{t}(w)) = P(\mathbf{y}(w))$.

Similarly as (89), using (50) and (51), we obtain

$$\Delta_3 \leq e^{-n\eta}, \quad (91)$$

$$\Delta_4 \leq e^{-n\eta}. \quad (92)$$

Invoking the definition of $\mathcal{A}_4(w)$ in (52), we conclude that

$$\Delta_5 \leq \sum_{w=1}^M \sum_{\mathbf{t} \in \mathcal{A}_4^c(w)} P(\mathbf{t}) \quad (93)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{\mathbf{t}(w): P(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}) \leq \\ e^{-n\eta} Q(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}, \hat{\mathbf{x}})}} P(\mathbf{t}(w)) \quad (94)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{\mathbf{t}(w): P(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}) \leq \\ e^{-n\eta} Q(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}, \hat{\mathbf{x}})}} P(s(w), \mathbf{z}) P(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}) P(\hat{\mathbf{x}}|s(w), \mathbf{z}) \quad (95)$$

$$\leq \sum_{w=1}^M \frac{1}{M} \sum_{\mathbf{t}(w)} P(s(w), \mathbf{z}) P(\hat{\mathbf{x}}|s(w), \mathbf{z}) e^{-n\eta} Q(\mathbf{x}(w), \mathbf{y}(w)|s(w), \mathbf{z}, \hat{\mathbf{x}}) \quad (96)$$

$$\leq e^{-n\eta}, \quad (97)$$

where (93) follows similarly as (85); (95) follows due to the Markov chain $(X^n(W), Y^n(W)) - (S(W), Z^n) - \hat{X}^n$.

Then, invoking (53), we upper bound Δ_6 as follows:

$$\Delta_6 = \sum_{w=1}^M \sum_{\substack{\mathbf{t} \in \mathcal{A}_5^c(w) \cap \mathcal{A}_6(w) \\ z^n \in \mathcal{D}(s^M, w), d(\mathbf{x}(w), \hat{\mathbf{x}}) \leq D}} P(\mathbf{t}) \quad (98)$$

$$\leq \sum_{w=1}^M \frac{1}{M} \sum_{\substack{\mathbf{t}(w): P(\mathbf{y}(w), z^n) \leq \\ Q(\mathbf{y}(w), z^n | s(w)) e^{-n(\eta + R^c)}}} P(\mathbf{t}(w)) \quad (99)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{(\mathbf{y}(w), s(w), \mathbf{z}): P(\mathbf{y}(w), \mathbf{z}) \\ \leq Q(\mathbf{y}(w), z^n | s(w)) e^{-n(\eta + R^c)}}} P(\mathbf{y}(w), z^n) P(s(w) | \mathbf{y}(w)) \sum_{\hat{\mathbf{x}}} P(\hat{\mathbf{x}} | s(w), \mathbf{z}) \quad (100)$$

$$\leq \sum_{w=1}^M \frac{1}{M} \sum_{\mathbf{y}(w), s(w), \mathbf{z}} Q(\mathbf{y}(w), z^n | s(w)) e^{-n(\eta + R^c)} \quad (101)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{s(w)} e^{-n(\eta + R^c)} \quad (102)$$

$$\leq e^{-n\eta}, \quad (103)$$

where (99) follows from ignoring the constraint that $(\mathbf{x}^M, \mathbf{y}^M, s^M, \mathbf{z}, \hat{\mathbf{x}}) \in \mathcal{A}_7(w)$, invoking the definition of $\mathcal{A}_4(w)$ in (52) and using (90); (101) follows since $P(s(w) | \mathbf{y}(w)) \leq 1$ for all $w \in \mathcal{M}$; (103) follows since $\sum_{s(w)} = |\mathcal{L}| = L$ for each $w \in \mathcal{M}$ and the fact that $L \leq e^{nR^c}$ from (65).

Finally, invoking (54), we upper bound Δ_7 as follows:

$$\Delta_7 \leq \sum_{w=1}^M \sum_{\substack{\mathbf{t} \in \mathcal{A}_6^c(w): \\ z^n \in \mathcal{D}(s^M, w)}} P(\mathbf{t}) \quad (104)$$

$$= \sum_{w=1}^M \sum_{\substack{\mathbf{t}: \mathbf{z} \in \mathcal{D}(s^M, w) \\ P(\mathbf{z} | s(w)) \leq e^{nR^i} e^{-n\eta} Q(\mathbf{z})}} P(\mathbf{t}) \quad (105)$$

$$= \sum_{w=1}^M \frac{1}{M} \sum_{\substack{s^M, \mathbf{z}: \mathbf{z} \in \mathcal{D}(s^M, w) \\ P(\mathbf{z} | s(w)) \leq e^{nR^i} e^{-n\eta} Q(\mathbf{z})}} P(s^M) P(\mathbf{z} | s(w)) \quad (106)$$

$$\leq \sum_{w=1}^M \frac{1}{M} \sum_{s^M, \mathbf{z}: \mathbf{z} \in \mathcal{D}(s^M, w) P(\mathbf{z} | s(w)) \leq e^{nR^i} e^{-n\eta} Q(\mathbf{z})} P(s^M) e^{nR^i} e^{-n\eta} Q(\mathbf{z}) \quad (107)$$

$$\leq e^{-n\eta} \frac{e^{nR^i}}{M} \sum_{w=1}^M \sum_{s^M} \sum_{z^n: z^n \in \mathcal{D}(s^M, w)} P(s^M) Q(z^n) \quad (108)$$

$$\leq e^{-n\eta} \sum_{w=1}^M \sum_{s^M} P(s^M) Q(\mathcal{D}(s^M, w)) \quad (109)$$

$$= e^{-n\eta} \sum_{s^M} P(s^M) Q\left(\bigcup_{w=1}^M \mathcal{D}(s^M, w)\right). \quad (110)$$

$$\leq e^{-n\eta} \quad (111)$$

where (104) follows from dropping the constraint that $(\mathbf{x}^M, \mathbf{y}^M, s^M, \mathbf{z}, \hat{\mathbf{x}}) \in \mathcal{A}_7(w)$; (105) follows invoking the definition of $\mathcal{A}_6(w)$ in (54); (106) follows since for each $w \in \mathcal{M}$,

$$\prod_{i \neq w, i \in \mathcal{M}} \sum_{\mathbf{x}(i), \mathbf{y}(i)} P(\mathbf{x}(i), \mathbf{y}(i), s(i)) = \prod_{i \neq w, i \in \mathcal{M}} P_S(s(i)), \quad (112)$$

$$\sum_{\mathbf{x}(w), \mathbf{y}(w)} P(x(w), y(w), s(w)) P(\mathbf{z}|\mathbf{x}(w)) = P(s(w), \mathbf{z}) = P(s(w)) P(\mathbf{z}|s(w)), \quad (113)$$

$$P(s^M) = \prod_{i \in \mathcal{M}} P(s(i)); \quad (114)$$

(108) follows since $M \geq e^{nR^i}$ due to (66); (110) follows since that decoding regions are disjoint for different $w \in \mathcal{M}$.

The proof of Lemma 9 is complete by combining (82), (89), (91), (92), (97), (103) and (111).

B. Proof of Lemma 10

Recall that in Section IV-A, we choose $U_i(W) = (S(W), Y^{i-1}(W), Z_{i+1}^n)$ and $V_i(W) = (S(W), Z_{i+1}^n)$. Then, in the following, let $U_i = (S, Y^{i-1}, Z_{i+1}^n)$ and $V_i = (S, Z_{i+1}^n)$. Recall that for $i = 1, \dots, n$, $Q_{X_i Y_i Z_i U_i \hat{X}_i}$ is any generic distribution and $Q_{Y_i}, Q_{Z_i}, Q_{X_i|Y_i Z_i U_i}, Q_{X_i Y_i|Z_i U_i \hat{X}_i}, Q_{Y_i Z_i|U_i}$ are induced by $Q_{X_i Y_i Z_i U_i \hat{X}_i}$. Further, note that in Lemma 9, we are free to choose the distributions $Q_{Y^n}, Q_{Z^n}, Q_{Z^n|SY^n}, Q_{X^n|SY^n Z^n}, Q_{X^n Y^n|SZ^n \hat{X}^n}, Q_{Y^n Z^n|S}$. Our choices for these distributions are as follows:

$$Q_{Y^n}(y^n) := \prod_{i=1}^n Q_{Y_i}(y_i), \quad (115)$$

$$Q_{Z^n}(z^n) := \prod_{i=1}^n Q_{Z_i}(z_i), \quad (116)$$

$$Q_{Z^n|SY^n}(z^n|s, y^n) := \prod_{i=1}^n Q_{Z_i|SY^i Z_{i+1}^n}(z_i|s, y^i, z_{i+1}^n) \quad (117)$$

$$= \prod_{i=1}^n Q_{Z_i|Y_i U_i}(z_i|y_i, u_i) \quad (118)$$

$$Q_{X^n|SY^n Z^n}(x^n|s, y^n, z^n) := \prod_{i=1}^n Q_{X_i|SY^i Z_i^n}(x_i|s, y^i, z_i^n) \quad (119)$$

$$= \prod_{i=1}^n Q_{X_i|Y_i Z_i U_i}(x_i|y_i, z_i, u_i) \quad (120)$$

$$Q_{X^n Y^n|SZ^n \hat{X}^n}(x^n, y^n|s, z^n, \hat{x}^n) := \prod_{i=1}^n Q_{X_i Y_i|SY^{i-1} Z_i^n \hat{X}^n}(x_i, y_i|s, y^{i-1}, z_i^n, \hat{x}_i) \quad (121)$$

$$= \prod_{i=1}^n Q_{X_i Y_i|Z_i U_i \hat{X}_i}(x_i, y_i|z_i, u_i, \hat{x}_i), \quad (122)$$

$$Q_{Y^n Z^n|S}(y^n, z^n|s) := \prod_{i=1}^n Q_{Y_i Z_i|SY^{i-1} Z_{i+1}^n}(y_i, z_i|s, y^{i-1}, z_{i+1}^n) \quad (123)$$

$$= \prod_{i=1}^n Q_{Y_i Z_i|U_i}(y_i, z_i|u_i). \quad (124)$$

Recall from Section IV-A that for each $w \in \mathcal{M}$, the joint distribution of $(X^n(W), Y^n(W), S(W), Z^n, \hat{X}^n)$ is the same and denoted as $P_{X^n Y^n S Z^n \hat{X}^n}$. The marginal distributions of $P_{X^n Y^n S Z^n \hat{X}^n}$ are as follows:

$$P_{Y^n}(y^n) = \prod_{i=1}^n P_{Y_i}(y_i), \quad (125)$$

$$P_{Z^n}(z^n) = \prod_{i=1}^n P_{Z_i}(z_i), \quad (126)$$

$$P_{Z^n|Y^n}(z^n|y^n) = \prod_{i=1}^n P_{Z_i|Y_i}(z_i|y_i), \quad (127)$$

$$P_{X^n|Y^n Z^n}(x^n|y^n, z^n) = \prod_{i=1}^n P_{X_i|Y_i Z_i}(x_i|y_i, z_i), \quad (128)$$

$$P_{X^n Y^n | S Z^n}(x^n, y^n | s, z^n) = \prod_{i=1}^n P_{X_i Y_i | S X^{i-1} Y^{i-1} Z^n}(x_i, y_i | s, x^{i-1}, y^{i-1}, z^n) \quad (129)$$

$$= \prod_{i=1}^n P_{X_i Y_i | S Y^{i-1} Z^n}(x_i, y_i | s, y^{i-1}, z_i^n) \quad (130)$$

$$= \prod_{i=1}^n P_{X_i Y_i | Z_i U_i}(x_i, y_i | z_i, u_i), \quad (131)$$

$$P_{Y^n Z^n}(y^n, z^n) = \prod_{i=1}^n P_{Y_i Z_i}(y_i, z_i), \quad (132)$$

$$P_{Z^n | S}(z^n | s) = \prod_{i=1}^n P_{Z_i | S Z_{i+1}^n}(z_i | s, z_{i+1}^n) \quad (133)$$

$$= \prod_{i=1}^n P_{Z_i | V_i}(z_i | v_i) \quad (134)$$

where (130) holds since the Markov chain $(X_i(W), Y_i(W)) - (S(W), Y^{i-1}(W), Z_i^n) - (X^{i-1}(W), Z^{i-1})$ holds. The proof of this Markov chain is similar as [20, Lemma 2] and thus omitted.

Recall the definitions of $\{\mathcal{B}_i(w)\}_{i=1}^7$ in (56) to (62). For each $w \in \mathcal{M}$, let

$$\tilde{\mathcal{B}}_5(w) := \left\{ \mathbf{t} : R^c \geq \frac{1}{n} \sum_{i=1}^n \log \frac{Q_{Y_i Z_i | U_i}(y_i(w), z_i | u_i(w))}{P_{Y_i Z_i}(y_i(w), z_i)} - \eta \right\}. \quad (135)$$

We remark that $\tilde{\mathcal{B}}_5(w)$ corresponds to $\mathcal{A}_5(w)$ (recall (53)) in Lemma 9 by applying the choice of $Q_{Y^n Z^n | S}$ in (124) and the definition in (132).

Recall the definition of $P_{W\mathbf{T}}$ in Section IV-A. Using Lemma 9 and (115)–(124) and (125)–(134), we obtain

$$P_c^{(n)}(f^{(n)}, g^{(n)}) \leq P_{W\mathbf{T}} \left\{ \bigcap_{i=1, i \neq 5}^7 \mathcal{B}_i(W) \cap \tilde{\mathcal{B}}_5(W) \right\} + 6e^{-n\eta}. \quad (136)$$

For each $w \in \mathcal{M}$, when $t \in \bigcap_{i=1, i \neq 5}^7 \mathcal{B}_i(w) \cap \tilde{\mathcal{B}}_5(w)$, invoking the constraints related with $\tilde{\mathcal{B}}_5(w)$ and $\mathcal{B}_6(w)$, we have that

$$R^c - R^i \geq \frac{1}{n} \sum_{i=1}^n \left(\log \frac{Q_{Y_i Z_i | U_i}(y_i(w), z_i | u_i(w))}{P_{Y_i Z_i}(y_i(w), z_i)} - \log \frac{P_{Z_i | V_i}(z_i | v_i(w))}{Q_{Z_i}(z_i)} \right) - 3\eta \quad (137)$$

$$= \frac{1}{n} \sum_{i=1}^n \left(\log \frac{Q_{Y_i Z_i | U_i}(y_i(w), z_i | u_i(w))}{P_{Y_i Z_i}(y_i(w), z_i)} \frac{Q_{Z_i}(z_i)}{P_{Z_i | V_i}(z_i | v_i(w))} \right) - 3\eta. \quad (138)$$

Hence, for each $w \in \mathcal{M}$, when $t \in \bigcap_{i=1, i \neq 5}^7 \mathcal{B}_i(w) \cap \tilde{\mathcal{B}}_5(w)$, we have $t \in \bigcap_{i=1, i}^7 \mathcal{B}_i(w)$ (recall (60)). Thus,

$$P_{W\mathbf{T}} \left\{ \bigcap_{i=1, i \neq 5}^7 \mathcal{B}_i(W) \cap \tilde{\mathcal{B}}_5(W) \right\} \leq P_{W\mathbf{T}} \left\{ \bigcap_{i=1}^7 \mathcal{B}_i(W) \right\}. \quad (139)$$

The proof of Lemma 10 is now complete.

C. Proof of Lemma 11

For each $w \in \mathcal{M}$ and any $(\alpha, \mu, \beta, \lambda) \in [0, 1]^3 \times \mathbb{R}_+$, define

$$\mathcal{F}_1(w) := \left\{ \mathbf{t} : 0 \geq \frac{\bar{\alpha}}{n} \sum_{i=1}^n \log \frac{Q_{Y_i}(y_i(w))}{P_{Y_i}(y_i(w))} - \bar{\alpha}\eta \right\}, \quad (140)$$

$$\mathcal{F}_2(w) := \left\{ \mathbf{t} : 0 \geq \frac{\bar{\alpha}}{n} \sum_{i=1}^n \log \frac{Q_{Y_i Z_i | U_i}(y_i(w), z_i | u_i(w))}{P_{Y_i Z_i}(y_i(w), z_i)} - \bar{\alpha}\eta \right\}, \quad (141)$$

$$\mathcal{F}_3(w) := \left\{ \mathbf{t} : 0 \geq \frac{\bar{\alpha}}{n} \sum_{i=1}^n \frac{Q_{X_i | Y_i Z_i U_i}(x_i(w) | y_i(w), z_i, u_i(w))}{P_{X_i | Y_i Z_i}(x_i(w) | y_i(w), z_i)} - \bar{\alpha}\eta, \right\}, \quad (142)$$

$$\mathcal{F}_4(w) := \left\{ \mathbf{t} : 0 \geq \frac{\bar{\alpha}}{n} \sum_{i=1}^n \frac{Q_{X_i Y_i | Z_i U_i \hat{X}_i}(x_i(w), y_i(w) | z_i, u_i(w), \hat{x}_i)}{P_{X_i Y_i | Z_i U_i}(x_i(w), y_i(w) | z_i, u_i(w))} - \bar{\alpha}\eta \right\}, \quad (143)$$

$$\mathcal{F}_5(w) := \left\{ \mathbf{t} : \alpha \bar{\mu} \bar{\beta} (R^c - R^i) \geq \alpha \bar{\mu} \bar{\beta} \frac{1}{n} \sum_{i=1}^n \log \left(\frac{Q_{Y_i Z_i | U_i}(y_i(w), z_i | u_i(w))}{P_{Y_i Z_i}(y_i(w), z_i)} \frac{Q_{Z_i}(z_i)}{P_{Z | V_i}(z_i | v_i(w))} \right) - 3\alpha \bar{\mu} \bar{\beta} \eta \right\}, \quad (144)$$

$$\mathcal{F}_6(w) := \left\{ \mathbf{t} : \alpha \bar{\mu} \beta R^i \leq \frac{\alpha \bar{\mu} \beta}{n} \sum_{i=1}^n \log \frac{P_{Z_i | V_i}(z_i | v_i(w))}{Q_Z(z_i)} + \alpha \bar{\mu} \beta \eta \right\}, \quad (145)$$

$$\mathcal{F}_7(w) := \left\{ \mathbf{t} : \alpha \mu D \geq \frac{\alpha \mu}{n} \sum_{i=1}^n \log e^{d(x_i(w), \hat{x}_i)} \right\}, \quad (146)$$

$$(147)$$

Invoking Lemma 10, for any $(\alpha, \mu, \beta, \lambda) \in [0, 1]^3 \times \mathbb{R}_+$, we obtain

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq P_{W\mathbf{T}} \left\{ \bigcap_{i=1}^7 \mathcal{F}_i(W) \right\} + 6e^{-n\eta}. \quad (148)$$

We make use of the following Cramér's bound for large deviations.

Lemma 13. *For any real valued random variable Z and any $\lambda > 0$, we have*

$$\Pr(Z \geq a) \leq \exp \left(-(\lambda a - \log \mathbb{E}[\exp(\lambda Z)]) \right). \quad (149)$$

Let

$$R(\alpha, \mu, \beta) := \alpha \bar{\mu} (\bar{\beta} (R^c - R^i) - \beta R^i) + \alpha \mu D, \quad (150)$$

$$c(\alpha, \mu, \beta) := 4\bar{\alpha} + 3\alpha \bar{\mu} \bar{\beta} + \alpha \bar{\mu} \beta \quad (151)$$

$$= 4\bar{\alpha} + \alpha \bar{\mu} (3 - 2\beta). \quad (152)$$

Recall the definition of $P_{W\mathbf{T}}$ in Section IV-A. In this subsection, whenever we use \Pr , we mean the probability with respect to $P_{W\mathbf{T}}$ unless otherwise stated. Recall the definition of $f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(\cdot)$ in (69). Combining (148), (150), (69) and Lemma 13, we obtain that for any $\lambda \in \mathbb{R}_+$,

$$\begin{aligned} & P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \\ & \leq \Pr \left\{ n \left(R(\alpha, \mu, \beta) + c(\alpha, \mu, \beta) \eta \right) \geq \sum_{i=1}^n \log f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(X_i(W), Y_i(W), Z_i, \hat{X}_i | U_i(W), V_i(W)) \right\} + 6e^{-n\eta} \end{aligned} \quad (153)$$

$$\begin{aligned} & = \Pr \left\{ - \sum_{i=1}^n \log f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(X_i(W), Y_i(W), Z_i, \hat{X}_i | U_i(W), V_i(W)) \geq -n \left(R(\alpha, \mu, \beta) + c(\alpha, \mu, \beta) \eta \right) \right\} + 6e^{-n\eta} \\ & \leq \exp \left\{ n \lambda \left(R(\alpha, \mu, \beta) + c(\alpha, \mu, \beta) \eta \right) \right\} \end{aligned} \quad (154)$$

$$+ \log \mathbb{E} \left[\exp \left(- \lambda \sum_{i=1}^n \log f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(X_i(W), Y_i(W), Z_i, \hat{X}_i | U_i(W), V_i(W)) \right) \right] \right\} + 6e^{-n\eta} \quad (155)$$

$$= \exp \left\{ n \left(\lambda R(\alpha, \mu, \beta) + \lambda c(\alpha, \mu, \beta) \eta - \frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) \right) \right\} + 6e^{-n\eta}, \quad (156)$$

and (156) follows from (70).

Choose η such that

$$\lambda R(\alpha, \mu, \beta) + \lambda c(\alpha, \mu, \beta) \eta - \frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) = -\eta. \quad (157)$$

Thus,

$$\eta = \frac{\frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) - \lambda R(\alpha, \mu, \beta)}{1 + \lambda c(\alpha, \mu, \beta)}. \quad (158)$$

Using the bound in (156), the definition in (158) and recalling (150) and (152), we obtain that

$$P_c^{(n)}(f^{(n)}, g^{(n)}, h^{(n)}, D) \leq 7 \exp(-n\eta) \quad (159)$$

$$\leq 7 \exp \left(-n \frac{\frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) - \lambda \alpha \bar{\mu} (\bar{\beta} R^c - R^i) - \lambda \alpha \mu D}{1 + \lambda (4\bar{\alpha} + \alpha \bar{\mu} (3 - 2\beta))} \right). \quad (160)$$

The proof of Lemma 11 is now complete.

D. Proof of lemma 12

1) *Removing Dependence on the Identification Index:* Recall from Section IV-A that for each w , the joint distribution of $(X^n(w), Y^n(w), S(w), Z^n, \hat{X}^n)$ is $P_{\mathbf{XYSZ}\hat{\mathbf{X}}}$ and $P_{X_i Y_i Z_i U_i V_i}$ is induced by $P_{\mathbf{XYSZ}\hat{\mathbf{X}}}$. Further, recall that Q_i denotes $Q_{X_i Y_i Z_i U_i \hat{X}_i}$ and P_i denotes $P_{X_i Y_i Z_i U_i V_i}$. Define

$$g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i) := \left(\frac{1}{f_{Q_i, P_i}^{(\alpha, \mu, \beta)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i)} \right)^\lambda. \quad (161)$$

Invoking (70), we obtain that

$$\exp \left(-\Omega^{(\alpha, \mu, \lambda)}(\{Q_i\}_{i=1}^n) \right) = \sum_{w=1}^M \frac{1}{M} \left(\sum_{x^n(w), y^n(w), s(w), z^n, \hat{x}^n} P_{\mathbf{XYSZ}\hat{\mathbf{X}}}(x^n(w), y^n(w), s(w), z^n, \hat{x}^n) \prod_{i=1}^n g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i(w), y_i(w), z_i, \hat{x}_i | u_i(w), v_i(w)) \right) \quad (162)$$

$$= \sum_{x^n, y^n, s, z^n, \hat{x}^n} P_{\mathbf{XYSZ}\hat{\mathbf{X}}}(x^n, y^n, s, z^n, \hat{x}^n) \prod_{i=1}^n g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i), \quad (163)$$

where $u_i = (s, y^{i-1}, z_{i+1}^n)$, $v_i = (s, z_{i+1}^n)$ and the joint distribution of $X^n, Y^n, S, Z^n, \hat{X}^n$ is $P_{\mathbf{XYSZ}\hat{\mathbf{X}}}$. Let $U_i = (S, Y^{i-1}, Z_{i+1}^n)$. Then we have the Markov chains $Z_i - X_i - Y_i - U_i$ and $(X_i, Y_i) - (U_i, Z_i) - \hat{X}_i$. In the following, all the distributions are induced by $P_{\mathbf{XYSZ}\hat{\mathbf{X}}}$ and we will drop the subscript of distributions for convenience.

2) *Preliminaries:* Invoking (163), we obtain that

$$\exp \left(-\Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) \right) = \sum_{s, z^n} P(s, z^n) \sum_{x^n, y^n, \hat{x}^n} P(x^n, y^n, \hat{x}^n | s, z^n) \prod_{i=1}^n g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i). \quad (164)$$

For $i = 1, \dots, n$, define

$$C_i(s, z^n) := \sum_{x^i, y^i, \hat{x}^i} P(x^i, y^i, \hat{x}^i | s, z^n) \prod_{j=1}^i g_{Q_j, P_j}^{(\alpha, \mu, \beta, \lambda)}(x_j, y_j, z_j, \hat{x}_j | u_j, v_j) \quad (165)$$

$$P^{(\alpha, \mu, \beta, \lambda)}(x^i, y^i, \hat{x}^i | s, z^n) := \frac{P(x^i, y^i, \hat{x}^i | s, z^n) \prod_{j=1}^i g_{Q_j, P_j}^{(\alpha, \mu, \beta, \lambda)}(x_j, y_j, z_j, \hat{x}_j | u_j, v_j)}{C_i(s, z^n)}, \quad (166)$$

$$\Phi_i^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_j\}_{j=1}^i) := C_i(s, z^n) / C_{i-1}(s, z^n). \quad (167)$$

Similarly as [20, Lemma 6], we obtain the following lemma.

Lemma 14. For $i = 1, \dots, n$ and any $(s, z^n, x^t, y^t, \hat{x}^t)$, we have

$$\Phi_i^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_j\}_{j=1}^i) = \sum_{x^i, y^i, \hat{x}^i} P^{(\alpha, \mu, \beta, \lambda)}(x^{i-1}, y^{i-1}, \hat{x}^{i-1} | s, z^n) P(x_i, y_i, \hat{x}_i | s, x^{i-1}, y^{i-1}, \hat{x}^{i-1}, z^n) g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i). \quad (168)$$

Hence, combining (164) and Lemma 14, we obtain that

$$\exp \left(-\Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) \right) = \sum_{s, z^n} P(s, z^n) \prod_{i=1}^n \Phi_i^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_j\}_{j=1}^i). \quad (169)$$

Then, for $i = 1, \dots, n$, define

$$\tilde{C}_i := \sum_{s, z^n} P(s, z^n) \prod_{j=1}^i \Phi_j^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_l\}_{l=1}^j), \quad (170)$$

$$P_{SZ^n}^{(\alpha, \mu, \beta, \lambda)|i}(s, z^n) := \frac{P(s, z^n) \prod_{j=1}^i \Phi_j^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_l\}_{l=1}^j)}{\tilde{C}_i}, \quad (171)$$

$$\Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i) := \tilde{C}_i / \tilde{C}_{i-1}. \quad (172)$$

Similarly as [20, Lemma 7], we obtain the following lemma.

Lemma 15. For $i = 1, \dots, n$, we have

$$\Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i) = \sum_{s, z^n} P_{SZ^n}^{(\alpha, \mu, \beta, \lambda)}(s, z^n) \Phi_i^{(\alpha, \mu, \beta, \lambda)}(s, z^n | \{Q_j\}_{j=1}^i). \quad (173)$$

Hence, invoking (169) and Lemma 15, we obtain

$$\exp\left(-\Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n)\right) = \prod_{i=1}^n \Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i). \quad (174)$$

3) *Final proof of Lemma 12:* Recall (19). Define

$$\hat{Q}_n := \left\{ Q_{XYZU\hat{X}} : |\mathcal{U}| \leq |\mathcal{L}||\mathcal{X}|^{n-1}|\mathcal{Y}|^{n-1}|\mathcal{Z}|^{n-1} \right\}, \quad (175)$$

$$\hat{\Omega}_n^{(\alpha, \mu, \beta, \lambda)} := \min_{Q_{XYZU\hat{X}} \in \hat{Q}_n} \Omega^{(\alpha, \mu, \beta, \lambda)}(Q_{XYZU\hat{X}}). \quad (176)$$

Recall that $u_i = (s, y^{i-1}, z_{i+1}^n)$. For each $i = 1, \dots, n$, define

$$\begin{aligned} & P^{(\alpha, \mu, \beta, \lambda)}(s, x_i, y^i, z_i^n, \hat{x}_i) \\ &= P^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, u_i, \hat{x}_i) \end{aligned} \quad (177)$$

$$:= \sum_{x^{i-1}, z^{i-1}, \hat{x}^{i-1}} P^{(\alpha, \mu, \beta, \lambda)}(s, z^n) P^{(\alpha, \mu, \beta, \lambda)}(x^{i-1}, y^{i-1}, \hat{x}^{i-1} | s, z^n) P(x_i, y_i, \hat{x}_i | s, x^{i-1}, y^{i-1}, \hat{x}^{i-1}, z^n), \quad (178)$$

where $P^{(\alpha, \mu, \beta, \lambda)}(x^{i-1}, y^{i-1}, \hat{x}^{i-1} | s, z^n)$ was defined in (166) and $P^{(\alpha, \mu, \beta, \lambda)}(s, z^n)$ was defined in (171).

Invoking Lemmas 14, 15 and (178), we obtain that for $i = 1, \dots, n$,

$$\Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i) = \sum_{x_i, y_i, z_i, u_i, \hat{x}_i} P^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, u_i, \hat{x}_i) g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i). \quad (179)$$

Note that $Q_i = Q_{X_i Y_i Z_i U_i \hat{X}_i}$ can be chosen arbitrarily for all $i = 1, \dots, n$. Here we apply the recursive method. For each $i = 1, \dots, n$, we choose $Q_{X_i Y_i Z_i U_i \hat{X}_i}$ such that

$$Q_{X_i Y_i Z_i U_i \hat{X}_i}(x_i, y_i, z_i, u_i, \hat{x}_i) = P^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, u_i, \hat{x}_i). \quad (180)$$

Then, let $Q_{Y_i}, Q_{Z_i}, Q_{X_i | Y_i Z_i U_i}, Q_{X_i Y_i | Z_i U_i \hat{X}_i}, Q_{Y_i Z_i | U_i}, Q_{Y_i Z_i}, Q_{X_i Y_i | Z_i U_i}, Q_{Z_i | U_i}$ be induced by $Q_{X_i Y_i Z_i U_i \hat{X}_i}$. Thus, we have $Q_{X_i Y_i Z_i U_i \hat{X}_i} \in \hat{Q}_n$.

Using the definition in (161), define

$$h_{Q_i, P_{X_i Y_i Z_i}}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i) := g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(x_i, y_i, z_i, \hat{x}_i | u_i, v_i) \left(\frac{P_{X_i Y_i | Z_i U_i}^{\lambda \bar{\alpha}}(x_i, y_i | z_i, u_i)}{Q_{X_i Y_i | Z_i U_i}^{\lambda \bar{\alpha}}(x_i, y_i | z_i, u_i)} \frac{P_{Z_i | V_i}^{\lambda \alpha \bar{\mu} \beta}(z_i | v_i)}{Q_{Z_i | U_i}^{\lambda \alpha \bar{\mu} \beta}(z_i | u_i)} \right)^{-1}. \quad (181)$$

In the following, for simplicity, we will drop the subscripts of the distributions. From (179), we obtain

$$\begin{aligned} \Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i) &= \mathbb{E}_{\{Q_j\}_{j=1}^i} [g_{Q_i, P_i}^{(\alpha, \mu, \beta, \lambda)}(X_i, Y_i, Z_i, \hat{X}_i | U_i, V_i)] \\ &= \mathbb{E}_{\{Q_j\}_{j=1}^i} \left[h_{Q_i, P_{X_i Y_i Z_i}}^{(\alpha, \mu, \beta, \lambda)}(X_i, Y_i, Z_i, \hat{X}_i | U_i) \frac{P^{\lambda \bar{\alpha}}(X_i, Y_i | Z_i, U_i)}{Q^{\lambda \bar{\alpha}}(X_i, Y_i | Z_i, U_i)} \frac{P^{\lambda \alpha \bar{\mu} \beta}(Z_i | V_i)}{Q^{\lambda \alpha \bar{\mu} \beta}(Z_i | U_i)} \right] \end{aligned} \quad (182)$$

$$\begin{aligned} &\leq \left(\mathbb{E}_{\{Q_j\}_{j=1}^i} \left[\left\{ h_{Q_i, P_{X_i Y_i Z_i}}^{(\alpha, \mu, \beta, \lambda)}(X_i, Y_i, Z_i, \hat{X}_i | U_i) \right\}^{\frac{1}{1 - \lambda \bar{\alpha} - \lambda \alpha \bar{\mu} \beta}} \right] \right)^{1 - \lambda \bar{\alpha} - \lambda \alpha \bar{\mu} \beta} \\ &\quad \times \left(\mathbb{E}_{\{Q_j\}_{j=1}^i} \left[\frac{P(X_i, Y_i | Z_i, U_i)}{Q(X_i, Y_i | Z_i, U_i)} \right] \right)^{\lambda \bar{\alpha}} \left(\mathbb{E}_{\{Q_j\}_{j=1}^i} \left[\frac{P(Z_i | V_i)}{Q(Z_i | U_i)} \right] \right)^{\lambda \alpha \bar{\mu} \beta} \end{aligned} \quad (183)$$

$$= \exp \left(- \left(1 - \lambda \bar{\alpha} - \lambda \alpha \bar{\mu} \beta \right) \Omega^{(\alpha, \mu, \beta, \frac{\lambda}{1 - \lambda \bar{\alpha} - \lambda \alpha \bar{\mu} \beta})}(Q_i) \right) \quad (184)$$

$$= \exp \left(- \frac{\Omega^{(\alpha, \mu, \beta, \theta)}(Q_i)}{1 + \theta \bar{\alpha} + \theta \alpha \bar{\mu} \beta} \right) \quad (185)$$

$$\leq \exp \left(- \frac{\hat{\Omega}_n^{(\alpha, \mu, \beta, \theta)}}{1 + \theta \bar{\alpha} + \theta \alpha \bar{\mu} \beta} \right) \quad (186)$$

$$= \exp \left(- \frac{\Omega^{(\alpha, \mu, \beta, \theta)}}{1 + \theta \bar{\alpha} + \theta \alpha \bar{\mu} \beta} \right), \quad (187)$$

where (182) follows from (181); (183) follows from Hölder's inequality; (184) follows from the definitions in (19) and (181); (185) follows from (73) and (74); (186) follows since $Q_{XYZU\hat{X}}^* \in \hat{\mathcal{Q}}_n$ (recall (175)); (187) follows from since the cardinality bound $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}||\mathcal{Z}||\hat{\mathcal{X}}|$ is sufficient to describe $\hat{\Omega}_n^{(\alpha, \beta, \mu, \lambda)}$ (the proof of this fact is similar to [20, Property 4 a]) and thus omitted).

Hence, combining (174) and (187), we obtain that

$$\frac{1}{n} \Omega^{(\alpha, \mu, \beta, \lambda)}(\{Q_i\}_{i=1}^n) = -\frac{1}{n} \sum_{i=1}^n \log \Lambda_i^{(\alpha, \mu, \beta, \lambda)}(\{Q_j\}_{j=1}^i) \quad (188)$$

$$\geq \frac{\Omega^{(\alpha, \mu, \beta, \theta)}}{1 + \theta \bar{\alpha} + \theta \alpha \bar{\mu} \beta}. \quad (189)$$

Finally, combining (72) and (189), we conclude that

$$\underline{\Omega}^{(\alpha, \mu, \beta, \lambda)} \geq \frac{\Omega^{(\alpha, \mu, \beta, \theta)}}{1 + \theta \bar{\alpha} + \theta \alpha \bar{\mu} \beta}. \quad (190)$$

The proof of Lemma 12 is now complete.

E. Proof of Lemma 2

Before proceeding the proof of Lemma 2, we present two alternative expressions of the rate-distortion region. Let

$$\mathcal{P} := \left\{ Q_{XYZU\hat{X}} : |\mathcal{U}| \leq |\mathcal{Y}| + 2, Z - X - Y - U, Q_X = P_X, Q_{Y|X} = P_{Y|X}, Q_{Z|X} = P_{Z|X}, \right. \\ \left. (X, Y) - (U, Z) - \hat{X} \right\}, \quad (191)$$

$$\mathcal{R}_{\text{ran}} := \bigcup_{Q_{XYZU\hat{X}} \in \mathcal{P}} \mathcal{R}(Q_{XYZU\hat{X}}). \quad (192)$$

Let

$$\mathcal{P}_{\text{sh}} := \left\{ Q_{XYZU\hat{X}} : |\mathcal{U}| \leq |\mathcal{Y}|, Z - X - Y - U, Q_X = P_X, Q_{Y|X} = P_{Y|X}, Q_{Z|X} = P_{Z|X}, \right. \\ \left. (X, Y) - (U, Z) - \hat{X} \right\} \quad (193)$$

Recall the definition of \mathcal{Q} in (17). Recall that given a number $a \in [0, 1]$, we define $\bar{a} = 1 - a$. Then for any $(\alpha, \mu, \beta) \in (0, 1] \times [0, 1]^2$, define

$$R^{(\mu, \beta)} := \min_{Q_{XYZU\hat{X}} \in \mathcal{P}_{\text{sh}}} \left\{ \bar{\mu} \bar{\beta} I(Q_{YZ}, Q_{U|YZ}) - \bar{\mu} I(Q_Z, Q_{Z|U}) + \mu \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})] \right\}, \quad (194)$$

$$\tilde{R}^{(\alpha, \mu, \beta)} := \min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \left\{ \bar{\alpha} \left(D(Q_Y \| P_Y) + D(Q_{Z|YU} \| P_{Z|Y} | Q_{YU}) + D(Q_{X|YZU} \| P_{X|YZ} | Q_{YZU}) \right) \right. \\ \left. + I(Q_{\hat{X}|ZU}, Q_{XY|ZU\hat{X}} | Q_{ZU}) \right\} + \alpha \left(\bar{\mu} \bar{\beta} I(Q_{YZ}, Q_{U|YZ}) - \bar{\mu} I(Q_Z, Q_{Z|U}) + \mu \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})] \right) \quad (195)$$

$$\mathcal{R}_{\text{sh}} := \bigcap_{(\mu, \beta) \in [0, 1]^2} \left\{ (R^i, R^c, D) : \bar{\mu} \bar{\beta} R^c - \bar{\mu} R^i + \mu D \geq R^{(\mu, \beta)} \right\}, \quad (196)$$

$$\tilde{\mathcal{R}}_{\text{sh}} := \bigcap_{(\alpha, \mu, \beta) \in (0, 1] \times [0, 1]^2} \left\{ (R^i, R^c, D) : \bar{\mu} \bar{\beta} R^c - \bar{\mu} R^i + \mu D \geq \frac{1}{\alpha} \tilde{R}^{(\alpha, \mu, \beta)} \right\}. \quad (197)$$

We then have the following lemma which presents two alternative expressions of the rate-distortion region.

Lemma 16. *The following conclusions are true.*

i) *Recalling the definition of \mathcal{R}^* in (15), we have*

$$\mathcal{R}_{\text{sh}} = \mathcal{R}_{\text{ran}} = \mathcal{R}^* = \mathcal{R}; \quad (198)$$

ii) *Let*

$$\alpha_0 := \frac{1}{1 + 8 \left(\log |\mathcal{Y}| |\mathcal{Z}|^2 e^{d^+} \right)}, \quad (199)$$

$$\xi := 2 \frac{\alpha}{\alpha} \log \left(|\mathcal{Y}| |\mathcal{Z}|^2 e^{d^+} \right), \quad (200)$$

$$c(\alpha, |\mathcal{Y}|, |\mathcal{Z}|) := \sqrt{\xi} \log \left(|\mathcal{U}|^4 |\mathcal{Y}|^2 |\mathcal{Z}|^2 \xi^{-3/2} e^{d^+} \right). \quad (201)$$

For any $\alpha \in (0, \alpha_0]$, we have

$$R^{(\mu, \beta)} - c(\alpha, |\mathcal{Y}|, |\mathcal{Z}|) \leq \frac{1}{\alpha} \tilde{R}^{(\alpha, \mu, \beta)} \leq R^{(\mu, \beta)}. \quad (202)$$

iii) We have

$$\tilde{\mathcal{R}}_{\text{sh}} = \mathcal{R}_{\text{sh}}. \quad (203)$$

The proof of Lemma 16 is similar to that of [20, Property 3] and thus omitted. We remark Lemma 16 plays an important role in the proof of Lemma 2.

In the following, we present the formal proof of Lemma 2.

1) *Proof of Conclusion i):* Let

$$Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, \theta)}(x, y, z, u, \hat{x}) := \frac{Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta)}(x, y, z, u, \hat{x}) \exp\left(-\theta \omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u)\right)}{\sum_{x, y, z, u, \hat{x}} Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta)}(x, y, z, u, \hat{x}) \exp\left(-\theta \omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u)\right)}. \quad (204)$$

Invoking (19), we obtain

$$\frac{\partial \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}})}{\partial \theta} = \sum_{x, y, z, u, \hat{x}} Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, \theta)}(x, y, z, u, \hat{x}) \omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u), \quad (205)$$

and

$$\frac{\partial^2 \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}})}{\partial \theta^2} = -\text{Var}_{Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, \theta)}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right]. \quad (206)$$

Combining (205) and (206) and applying a Taylor expansion to $\Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}})$ around $\theta = 0$, we obtain that for some $\tau \in (0, \theta] \subset (0, 1]$,

$$\Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}) = \Omega^{(\alpha, \mu, \beta, 0)}(Q_{XYZU\hat{X}}) + \theta \frac{\partial \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}})}{\partial \theta} + \frac{\tau^2}{2} \frac{\partial^2 \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}})}{\partial \theta^2} \quad (207)$$

$$= 0 + \theta \mathbb{E}_{Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, 0)}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right] - \frac{\tau^2}{2} \text{Var}_{Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, 0)}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right] \quad (208)$$

$$\geq \theta \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right] - \frac{\tau^2 \rho}{2}, \quad (209)$$

where (209) follows from the definition of ρ in (23) and the fact that $Q_{XYZU\hat{X}}^{(\alpha, \mu, \beta, 0)} = Q_{XYZU\hat{X}}$ (recall (204)).

Using (198) (conclusion i)) in Lemma 16, we obtain that if $(R^c + \tau, R^i - \tau, D + \tau) \notin \mathcal{R}$, there exists $(\mu^*, \beta^*) \in [0, 1]^2$ such that

$$\bar{\mu}^* \bar{\beta}^* (R^c + \tau) - \bar{\mu}^* (R^i - \tau) + \mu^* (D + \tau) < R^{(\mu^*, \beta^*)}. \quad (210)$$

Further, invoking (202) and (210), we obtain that for any $\alpha \in (0, \alpha_0)$ (recall (199)),

$$\bar{\mu}^* \bar{\beta}^* R^c - \bar{\mu}^* R^i + \mu^* D + \tau < \frac{1}{\alpha} \tilde{R}^{(\alpha, \mu^*, \beta^*)} + c(\alpha, |\mathcal{Y}|, |\mathcal{Z}|). \quad (211)$$

Fix $\delta > 0$ and let $\alpha = \tau^{2+\delta}$. Then (211) holds for any $\tau \in (0, \alpha_0^{1/(2+\delta)})$. Combining (200) and (201), we see that as $\tau \rightarrow 0$,

$$\xi = 2 \frac{\tau^{2+\delta}}{1 - \tau^{2+\delta}} \log(|\mathcal{Y}| |\mathcal{Z}| e^{d^+}) \rightarrow 0, \quad (212)$$

$$\frac{c(\alpha, |\mathcal{Y}|, |\mathcal{Z}|)}{\tau} = \sqrt{\frac{2 \log(|\mathcal{Y}| |\mathcal{Z}| e^{d^+})}{1 - \tau^{2+\delta}}} \tau^\delta \left(4 \log |\mathcal{U}| + 2 \log |\mathcal{Y}| + 2 \log |\mathcal{Z}| + d^+ - \frac{3}{2} \log \xi \right) \quad (213)$$

$$\leq \sqrt{\frac{2 \log(|\mathcal{Y}| |\mathcal{Z}| e^{d^+})}{1 - \tau^{2+\delta}}} \tau^\delta \left(4 \log |\mathcal{U}| + 2 \log |\mathcal{Y}| + 2 \log |\mathcal{Z}| + d^+ \right) + \frac{3}{2} h_b \left(\sqrt{\frac{2 \log(|\mathcal{Y}| |\mathcal{Z}| e^{d^+})}{1 - \tau^{2+\delta}}} \tau^\delta \right) \quad (214)$$

$$\rightarrow 0, \quad (215)$$

where $h_b(a) = -a \log a - (1-a) \log(1-a)$ is the binary entropy function; (215) holds since for $(a, b) \in [0, 1]$, $a \leq \frac{\sqrt{a}}{b}$, $-a \log(a) \leq h_b(a)$ and $h_b(a)$ is increasing in a for $a \in (0, \frac{1}{2})$. Hence, invoking (215), we conclude that there exists $\nu \leq \alpha_0^{1/(2+\delta)}$ such that for any $\tau \in (0, \nu]$,

$$c(\alpha, |\mathcal{Y}|, |\mathcal{Z}|) \leq \frac{\tau}{2}. \quad (216)$$

Referring to (211) and (216) and noting that $\alpha = \tau^{2+\delta}$, we obtain that

$$\bar{\mu}^* \bar{\beta}^* R^c - \bar{\mu}^* R^i + \mu^* D + \frac{\tau}{2} < \frac{1}{\tau^{2+\delta}} \tilde{R}^{(\tau^{2+\delta}, \mu^*, \beta^*)}. \quad (217)$$

Then, invoking (22), we conclude that for any $\tau \in (0, \nu]$,

$$F(R^i, R^c, D) \geq \sup_{\theta > 0} F^{(\tau^{2+\delta}, \mu^*, \beta^*, \theta)}(R^i, R^c, D) \quad (218)$$

$$= \sup_{\theta > 0} \frac{\Omega(\tau^{2+\delta}, \mu^*, \beta^*, \theta) - \theta \tau^{2+\delta} (\bar{\mu}^* (\bar{\beta}^* R^c - R^i) + \mu^* D)}{1 + 5\theta(1 - \tau^{2+\delta}) + \theta \tau^{2+\delta} \bar{\mu}^* (3 - \beta^*)} \quad (219)$$

$$\geq \sup_{\theta > 0} \frac{1}{1 + 8\theta} \left\{ -\frac{1}{2} \rho \theta^2 + \theta \tilde{R}^{(\tau^{2+\delta}, \mu^*, \beta^*)} - \theta \tau^{2+\delta} (\bar{\mu}^* (\bar{\beta}^* R^c - R^i) + \mu^* D) \right\} \quad (220)$$

$$\geq \sup_{\theta > 0} \frac{1}{1 + 8\theta} \left\{ -\frac{1}{2} \rho \theta^2 + \frac{\tau^{3+\delta} \theta}{2} \right\} \quad (221)$$

$$= \frac{\rho}{2} \phi^2 (\tau^{3+\delta} / 6\rho), \quad (222)$$

where (221) follows from (217). In the following, we explain in detail why (220) and (222) holds.

(220) holds for the following three reasons:

i) Recalling $(\mu^*, \beta^*) \in [0, 1]^2$ and $\tau^{2+\delta} < \alpha_0 < 1$ (recall (199)), we have

$$1 + 4\theta(1 - \tau^{2+\delta}) + \theta \tau^{2+\delta} \bar{\mu}^* (3 - \beta^*) \leq 1 + 4\theta + 3\theta \quad (223)$$

$$= 1 + 7\theta; \quad (224)$$

ii) Referring to (20) and (209), we obtain

$$\Omega^{(\tau^{2+\delta}, \mu^*, \beta^*, \theta)} = \min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \Omega^{(\tau^{2+\delta}, \mu^*, \beta^*, \theta)}(Q_{XYZU\hat{X}}) \quad (225)$$

$$\geq \min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \theta \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\tau^{2+\delta}, \mu^*, \beta^*)}(X, Y, Z, \hat{X}|U) \right] - \frac{\tau^2 \rho}{2}; \quad (226)$$

iii) Referring to (18), we obtain that for any $(\alpha, \mu, \beta) \in [0, 1]^3$, we have

$$\begin{aligned} & \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(X, Y, Z, \hat{X}|U) \right] \\ &= \bar{\alpha} \left(D(Q_Y \| P_Y) + D(Q_Z | YU \| P_Z | Y) Q_{YU} + D(Q_X | YZU \| P_X | YZ | P_Y ZU) + I(Q_{XY|ZU}, Q_{\hat{X}|XYZU} | Q_{ZU}) \right) \\ & \quad + \alpha \bar{\mu} \left(\bar{\beta} I(Q_{YZ}, Q_{U|YZ}) + D(Q_{YZ} \| P_{YZ}) + I(Q_Z, Q_{U|Z}) \right) + \alpha \mu \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})] \end{aligned} \quad (227)$$

$$\begin{aligned} & \geq \bar{\alpha} \left(D(Q_Y \| P_Y) + D(Q_Z | YU \| P_Z | Y) Q_{YU} + D(Q_X | YZU \| P_X | YZ | P_Y ZU) + I(Q_{XY|ZU}, Q_{\hat{X}|XYZU} | Q_{ZU}) \right) \\ & \quad + \alpha \bar{\mu} \left(\bar{\beta} I(Q_{YZ}, Q_{U|YZ}) + I(Q_Z, Q_{U|Z}) \right) + \alpha \mu \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})]. \end{aligned} \quad (228)$$

Combining (195) and (228), we obtain

$$\min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\tau^{2+\delta}, \mu^*, \beta^*)}(X, Y, Z, \hat{X}|U) \right] \geq \tilde{R}^{(\tau^{2+\delta}, \mu^*, \beta^*)}. \quad (229)$$

The bound in (222) then follows from basic calculations. Given (l, x, y) , let

$$s(l, x, y) := \frac{1}{1 + 8l} \frac{lx - l^2 y}{2}. \quad (230)$$

Then we have

$$\frac{\partial s(l, x, y)}{\partial l} = -\frac{8l^2 y + 2ly - x}{2(1 + 8l)^2} \quad (231)$$

$$\frac{\partial s(l, x, y)}{\partial l} = -\frac{8x + y}{(1 + 8l)^3}. \quad (232)$$

Thus, $s(l, x, y)$ is concave in l when $x > 0, y > 0$. Hence, the supremum is achieved by l^* such that

$$8(l^*)^2 y + 2l^* y - x = 0. \quad (233)$$

Thus, $l^* = \phi(\frac{x}{y})$ and $l^*x = (l^*)^2(2y + 8l^*y)$. Therefore,

$$\sup_{l>0} s(l, x, y) = \frac{l^*x - (l^*)^2y}{2(1 + 8l^*)} \quad (234)$$

$$= \frac{(l^*)^2(y + 8l^*y)}{2(1 + 8l^*)} \quad (235)$$

$$= \frac{(l^*)^2y}{2} \quad (236)$$

$$= \frac{y}{2}\phi^2\left(\frac{x}{y}\right). \quad (237)$$

Replacing (l, x, y) with $(\theta, \tau^{3+\delta}, \rho)$, we have (222).

2) *Proof of Conclusion ii):* If $(R^i, R^c, D) \in \mathcal{R} = \mathcal{R}_{\text{ran}}$, then there exists a joint distribution $Q_{XYZU\hat{X}}^* \in \mathcal{P}$ such that

$$R^c - R^i \geq I(Q_{U|Z}^*, Q_{Y|UZ}^* | Q_Z^*), \quad (238)$$

$$R^i \leq I(Q_Z^*, Q_{U|Z}^*), \quad (239)$$

$$D \geq \mathbb{E}_{Q_{X\hat{X}}^*} [d(X, \hat{X})]. \quad (240)$$

Hence, for any (α, μ, β) , we have

$$\bar{\mu}\bar{\beta}R^c - \bar{\mu}R^i + \mu D = \bar{\mu}\bar{\beta}(R^c - R^i) - \bar{\mu}\beta R^i + \mu D \quad (241)$$

$$\geq \bar{\mu}\bar{\beta}\left(I(Q_{YZ}^*, Q_{U|YZ}^*) - I(Q_Z^*, Q_{U|Z}^*)\right) - \bar{\mu}\beta I(Q_Z^*, Q_{U|Z}^*) + \mu \mathbb{E}_{Q_{X\hat{X}}^*} [d(X, \hat{X})] \quad (242)$$

$$= \bar{\mu}\bar{\beta}I(Q_{YZ}^*, Q_{U|YZ}^*) - \bar{\mu}I(Q_Z^*, Q_{U|Z}^*) + \mu \mathbb{E}_{Q_{X\hat{X}}^*} [d(X, \hat{X})]. \quad (243)$$

Using (209), we obtain

$$\Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}) \leq \theta \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u) \right]. \quad (244)$$

Combining (191) and (17), we conclude that $\mathcal{Q} \supseteq \mathcal{P}$. Thus, invoking (20) and (244), we have

$$\Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}) = \min_{Q_{XYZU\hat{X}} \in \mathcal{Q}} \Omega^{(\alpha, \mu, \beta, \theta)}(Q_{XYZU\hat{X}}) \quad (245)$$

$$\leq \min_{Q_{XYZU\hat{X}} \in \mathcal{P}} \theta \mathbb{E}_{Q_{XYZU\hat{X}}} \left[\omega_{Q_{XYZU\hat{X}}}^{(\alpha, \mu, \beta)}(x, y, z, \hat{x}|u) \right] \quad (246)$$

$$= \min_{Q_{XYZU\hat{X}} \in \mathcal{P}} \theta \alpha \left(\bar{\mu}\bar{\beta}I(Q_{YZ}, Q_{U|YZ}) - \bar{\mu}I(Q_Z, Q_{U|Z}) + \mu \mathbb{E}_{Q_{X\hat{X}}} [d(X, \hat{X})] \right) \quad (247)$$

$$\leq \theta \alpha \left(\bar{\mu}\bar{\beta}I(Q_{YZ}^*, Q_{U|YZ}^*) - \bar{\mu}I(Q_Z^*, Q_{U|Z}^*) + \mu \mathbb{E}_{Q_{X\hat{X}}^*} [d(X, \hat{X})] \right) \quad (248)$$

$$\leq \theta \alpha \left(\bar{\mu}\bar{\beta}R^c - \bar{\mu}R^i + \mu D \right). \quad (249)$$

Thus, combining (21) and (249), we obtain that

$$F^{(\alpha, \mu, \beta, \theta)} = \frac{\Omega^{(\alpha, \mu, \beta, \theta)} - \theta \alpha \left(\bar{\mu}(\bar{\beta}R^c - R^i) + \mu D \right)}{1 + 5\theta\bar{\alpha} + \theta\alpha\bar{\mu}(3 - \beta)} \quad (250)$$

$$\leq 0. \quad (251)$$

On the other hand, note that

$$\lim_{\theta \rightarrow 0} F^{(\alpha, \mu, \beta, \theta)} = 0. \quad (252)$$

Hence, combining (251) and (252), we conclude that

$$F = \sup_{(\alpha, \mu, \beta, \theta) \in [0, 1]^3 \times \mathbb{R}_+} F^{(\alpha, \mu, \beta, \theta)} = 0. \quad (253)$$

F. Proof of Extensions for the Biometrical Identification Problem

1) *Exponent of the Probability of Correct Decoding*: Specializing Lemma 9 to the biometrical problem (using $\mathcal{A}_5(w)$ only), we obtain that for any decoding function $g^{(n)}$ and any $\eta \geq 0$,

$$P_c^{(n)}(g^n) \leq \Pr \left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq R^i - \eta \right) + \exp(-n\eta). \quad (254)$$

Further, adopting the one-shot technique in [41], we conclude that there exists a decoding function $g^{(n)}$ and $\gamma \geq 0$ such that

$$P_c^{(n)}(g^n) \geq \frac{1}{1 + \exp(-n\gamma)} \Pr \left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq R^i + \gamma \right). \quad (255)$$

Due to the memoryless of the source and channel, we have that (X_i, Y_i, Z_i) is an i.i.d. sequence, distributed according to $P_X \times P_{Y|X} \times P_{Z|X}$. Specializing (255) with $\gamma = 0$ and using Cramér's theorem [50, Theorem 2.2.3], we obtain

$$P_c^{(n)}(g^n) \geq \frac{1}{2} \exp \left(-n \sup_{\lambda > 0} \left\{ \lambda R^i - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Y|Z}(Y|Z)}{P_Z(Z)} \right) \right] \right\} \right) \quad (256)$$

$$= \frac{1}{2} \exp \left(-n \sup_{\lambda > 0} \left\{ \lambda R^i - \log \mathbb{E} \left[\log \frac{P_{Z|Y}^\lambda(Z|Y)}{P_Z^\lambda(Z)} \right] \right\} \right). \quad (257)$$

Combining (254) and Lemma 13,

$$P_c^{(n)}(g^n) \leq \exp \left(-n\lambda(R^i - \eta) + \log \mathbb{E} \left[\exp \left(\lambda \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \right) \right] \right) + \exp(-n\eta) \quad (258)$$

$$= \exp \left(-n\lambda(R^i - \eta) + n \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right] \right) + \exp(-n\eta), \quad (259)$$

where (259) follows since (X_i, Y_i, Z_i) is an i.i.d. sequence.

Choose η such that

$$\eta = \lambda(R^i - \eta) - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right]. \quad (260)$$

In other words,

$$\eta = \frac{\lambda R^i - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right]}{1 + \lambda}. \quad (261)$$

With this choice of η , we obtain that

$$P_c^{(n)}(g^n) \leq 2 \exp(-n\eta) \quad (262)$$

$$= 2 \exp \left(-n \sup_{\lambda > 0} \frac{\lambda R^i - \log \mathbb{E} \left[\exp \left(\lambda \log \frac{P_{Z|Y}(Z|Y)}{P_Z(Z)} \right) \right]}{1 + \lambda} \right). \quad (263)$$

2) *Moderate Deviations Constant in the Strong Converse Regime* (42): Let

$$nR^i = \log M := nI(P_Y, P_{Z|Y}) + n\xi_n. \quad (264)$$

Invoking (255) and choosing $\gamma = \zeta \xi_n$ for some $\zeta > 0$, we conclude that there exists a sequence of decoding function $g^{(n)}$ such that

$$P_c^{(n)}(g^n) \geq \frac{1}{1 + \exp(-n\zeta \xi_n)} \Pr \left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq I(P_Y, P_{Z|Y}) + (1 + \zeta)\xi_n \right). \quad (265)$$

Invoking the moderate deviations principle [50, Theorem 3.7.1], we obtain that

$$\lim_{n \rightarrow \infty} - \frac{\log \Pr \left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq I(P_Y, P_{Z|Y}) + (1 + \zeta)\xi_n \right)}{n\xi_n^2} = \frac{(1 + \zeta)^2}{2V}. \quad (266)$$

Therefore,

$$\liminf_{n \rightarrow \infty} \frac{-\log P_c^{(n)}(g^{(n)})}{n\xi_n^2} \leq \limsup_{n \rightarrow \infty} \frac{\log(1 + \exp(-n\zeta\xi_n))}{n\xi_n^2} - \frac{\log \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq R^i + (1 + \zeta)\xi_n\right)}{n\xi_n^2} \quad (267)$$

$$= \frac{(1 + \zeta)^2}{2V}. \quad (268)$$

On the other hand, for any decoding function and any M such that

$$nR^i = \log M = I(P_Y, P_{Z|Y}) + n\xi_n, \quad (269)$$

invoking (254) and choosing $\eta = (1 + \zeta)\xi_n$, we obtain that

$$P_c^{(n)}(g^n) \leq \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq I(P_Y, P_{Z|Y}) + (1 - \zeta)\xi_n\right) + \exp(-n\zeta\xi_n). \quad (270)$$

Similar as (266), we conclude that the first term in (270) is of the order $\exp(-n\xi_n^2 \frac{(1-\zeta)^2}{2V})$ and thus dominates the right hand side of (270) for sufficiently large n . Hence,

$$\limsup_{n \rightarrow \infty} \frac{-\log P_c^{(n)}(g^{(n)})}{n\xi_n^2} \geq \frac{(1 - \zeta)^2}{2V}. \quad (271)$$

The proof is complete by letting $\zeta \rightarrow 0$.

3) *Moderate Deviations Constant* (43): Invoking (254), we obtain that for any decoding function $g^{(n)}$, we have

$$P_e^{(n)}(g^{(n)}) = 1 - P_c^{(n)}(g^{(n)}) \quad (272)$$

$$\geq \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} < R^i - \eta\right) - \exp(-n\eta). \quad (273)$$

Invoking (255), we obtain that there exists a decoding function $g^{(n)}$ such that

$$P_e^{(n)}(g^{(n)}) \leq 1 - \frac{1}{1 + \exp(-n\gamma)} \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq R^i + \gamma\right) \quad (274)$$

$$\leq \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} < R^i + \gamma\right) + \left(1 - \frac{1}{1 + \exp(-n\gamma)}\right) \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} \geq R^i + \gamma\right) \quad (275)$$

$$\leq \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} < R^i + \gamma\right) + \left(1 - \frac{1}{1 + \exp(-n\gamma)}\right) \quad (276)$$

$$\leq \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{P_{Z|Y}(Z_i|Y_i)}{P_Z(Z_i)} < R^i + \gamma\right) + \exp(-n\gamma). \quad (277)$$

The rest of the proof is similar to that in Appendix F2 by invoking (273), (277) with properly chosen value of γ and η as well as applying the moderate deviations principle.

4) *Second-order Asymptotics*: The result in Theorem 8 follows by i) letting $\gamma = \eta = \frac{\log n}{n}$ and ii) applying the Berry-Esseen theorem to (254) and (255) or to (273) and (277).

REFERENCES

- [1] (2016). [Online]. Available: [https://en.wikipedia.org/wiki/Shazam_\(service\)](https://en.wikipedia.org/wiki/Shazam_(service))
- [2] E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, 2014.
- [3] R. Naini and P. Moulin, "Fingerprint information maximization for content identification," in *IEEE ICASSP*, 2014, pp. 3809–3813.
- [4] H. Yu, P. Moulin, and S. Roy, "RGB-D video content identification," in *IEEE ICASSP*, 2013, pp. 3776–3780.
- [5] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, 2009.
- [6] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *IEEE ISIT*, 2003, pp. 82–82.
- [7] G. Dasarthy and S. C. Draper, "Upper and lower bounds on the reliability of content identification," in *IEEE IZS*, 2014, p. 100.
- [8] N. Merhav, "Reliability of universal decoding based on vector-quantized codewords," *arXiv:1609.08868*, 2016.
- [9] V. Yachongka and H. Yagi, "Reliability function and strong converse of biometrical identification systems," in *IEEE ISITA*, 2016.
- [10] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, 1973.
- [11] R. Naini and P. Moulin, "Model-based decoding metrics for content identification," in *IEEE ICASSP*, 2012, pp. 1829–1832.
- [12] H. Yu and P. Moulin, "Regularized adaboost for content identification," in *IEEE ICASSP*, 2013, pp. 3078–3082.
- [13] P. Moulin, "Statistical modeling and analysis of content identification," in *IEEE ITA*, 2010, pp. 1–5.

- [14] F. Farhadzadeh, S. Voloshynovskiy, O. Koval, and F. Beekhof, "Information-theoretic analysis of content based identification for correlated data," in *IEEE ITW*, 2011, pp. 205–209.
- [15] D. Gündüz, E. Tuncel, A. Goldsmith, and H. V. Poor, "Identification over multiple databases," in *IEEE ISIT*, 2009, pp. 2311–2315.
- [16] F. Farhadzadeh, K. Sun, and S. Fredowski, "Efficient two stage decoding scheme to achieve content identification capacity," in *IEEE ISIT*, 2014, pp. 3814–3818.
- [17] E. Tuncel, "Recognition capacity versus search speed in noisy databases," in *IEEE ISIT*, 2012, pp. 2566–2570.
- [18] Y. Oohama, "Exponent function for one helper source coding problem at rates outside the rate region," *arXiv:1504.05891*, 2015.
- [19] —, "New strong converse for asymmetric broadcast channels," *arXiv:1604.02901*, 2016.
- [20] —, "Exponent function for source coding with side information at the decoder at rates below the rate distortion function," *arXiv:1601.05650*, 2016.
- [21] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.
- [22] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.
- [23] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, 1977.
- [24] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [25] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [26] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *IEEE ISIT*, 2009, pp. 2316–2320.
- [27] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, 2015.
- [28] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 197–199, 1974.
- [29] Y. Altuğ, A. B. Wagner, and I. Kontoyiannis, "Lossless compression with moderate error probability," in *IEEE ISIT*, 2013, pp. 1744–1748.
- [30] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *Proc. 48th Annu. Allerton Conf.*, 2010, pp. 1334–1339.
- [31] C. T. Chubb, V. Y. Tan, and M. Tomamichel, "Moderate deviation analysis for classical communication over quantum channels," *arXiv:1701.03114*, 2017.
- [32] V. Y. F. Tan, "Moderate-deviations of lossy source coding for discrete and Gaussian sources," in *IEEE ISIT*, 2012, pp. 920–924.
- [33] V. Y. F. Tan, S. Watanabe, and M. Hayashi, "Moderate deviations for joint source-channel coding of systems with Markovian memory," in *IEEE ISIT*, 2014, pp. 1687–1691.
- [34] L. Zhou, V. Y. F. Tan, and M. Motani, "Second-order and moderate deviation asymptotics for successive refinement," *accepted to IEEE Trans. Inf. Theory*, 2017.
- [35] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," in *Trans. Third Prague Conf. Information Theory*, 1962, pp. 689–723.
- [36] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.
- [37] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [38] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1–2, pp. 1–184, 2014.
- [39] Y. Oohama *et al.*, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, 1994.
- [40] W. Kang, D. Cao, and N. Liu, "Deception with side information in biometric authentication systems," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1344–1350, 2015.
- [41] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *IEEE ISIT*, 2013, pp. 1287–1291.
- [42] J. Scarlett, "On the dispersions of the Gel'fand-Pinsker channel and dirty paper coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4569–4586, 2015.
- [43] S. Watanabe, "Second-order region for Gray-Wyner network," *IEEE Trans. Inf. Theory*, 2017.
- [44] L. Zhou, V. Y. F. Tan, and M. Motani, "Discrete lossy Gray-Wyner revisited: Second-order asymptotics, large and moderate deviations," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, pp. 1–1, 2016.
- [45] F. Willems, T. Kalker, J. Goseling, and J. P. Linnartz, "On the capacity of a biometrical identification system," in *IEEE ISIT*, 2003.
- [46] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, 2014.
- [47] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp-Lieb inequalities and strong converses for common randomness generation," in *IEEE ISIT*, 2016, pp. 1043–1047.
- [48] S. L. Fong and V. Y. F. Tan, "A proof of the strong converse theorem for Gaussian multiple access channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4376–4394, 2016.
- [49] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for Gaussian broadcast channels via the Gaussian poincare inequality," *arXiv:1509.01380*, 2015.
- [50] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer Science & Business Media, 2009, vol. 38.